



THE COGNITIVE AND NEURAL ORGANISATION OF SPEECH PROCESSING

EDITED BY: Patti Adank, Carolyn McGettigan and Sonja A. E. Kotz
PUBLISHED IN: Frontiers in Human Neuroscience



frontiers Research Topics



frontiers

Frontiers Copyright Statement

© Copyright 2007-2015 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-775-0

DOI 10.3389/978-2-88919-775-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

THE COGNITIVE AND NEURAL ORGANISATION OF SPEECH PROCESSING

Topic Editors:

Patti Adank, University College London, UK

Carolyn McGettigan, Royal Holloway University of London, UK

Sonja A. E. Kotz, Max Planck Institute Leipzig, Germany



Sequential sagittal gradient-echo MRI images of the human head and vocal tract (ordered left to right) were collected at a sample rate of 8 frames per second while a native speaker of British English (Ulster) produced the monosyllables /bead/, /bird/ and /bood/.
Image by Carolyn McGettigan.

Speech production and perception are two of the most complex actions humans perform. The processing of speech is studied across various fields and using a wide variety of research approaches. These fields include, but are not limited to, (socio)linguistics, phonetics, cognitive psychology, neurophysiology, and cognitive neuroscience. Research approaches range from behavioural studies to neuroimaging techniques such as Magnetoencephalography, electroencephalography (MEG/EEG) and functional Magnetic Resonance Imaging (fMRI), as well as neurophysiological approaches, such as the recording of Motor Evoked Potentials (MEPs), and Transcranial Magnetic Stimulation (TMS). Each of these approaches provides valuable information about specific aspects of speech processing. Behavioural testing can inform about the nature of the cognitive processes involved in speech processing, neuroimaging methods show

where (fMRI and MEG) in the brain these processes take place and/or elucidate on the time-course of activation of these brain areas (EEG and MEG), while neurophysiological methods (MEPs and TMS) can assess critical involvement of brain regions in the cognitive process. Yet, what is currently unclear is how speech researchers can combine methods such that a convergent approach adds to theory/model formulation, above and beyond the contribution of individual component methods? We expect that such combinations of approaches will significantly forward theoretical development in the field.

The present research topic comprise a collection of manuscripts discussing the cognitive and neural organisation of speech processing, including speech production and perception at the level of individual speech sounds, syllables, words, and sentences. Our goal was to use findings from a variety of disciplines, perspectives, and approaches to gain a more complete picture of the organisation of speech processing. The contributions are grouped around the following five main themes: 1) Spoken language comprehension under difficult listening conditions; 2) Sub-lexical processing; 3) Sensorimotor processing of speech; 4) Speech production. The contributions used a variety of research approaches, including behavioural experiments, fMRI, EEG, MEG, and TMS. Twelve of the 14 contributions were on speech perception processing, and the remaining two examined speech production. This Research Topic thus displays a wide variety of topics and research methods and this comprehensive approach allows an integrative understanding of currently available evidence as well as the identification of concrete venues for future research.

Citation: Adank, P., McGettigan, C., Kotz, S. A. E., eds. (2016). *The Cognitive and Neural Organisation of Speech Processing*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-775-0

Table of Contents

- 05 Editorial: Current research and emerging directions on the cognitive and neural organization of speech processing**
Patti Adank, Carolyn McGettigan and Sonja A. E. Kotz
- 07 Cortical entrainment to continuous speech: Functional roles and interpretations**
Nai Ding and Jonathan Z. Simon
- 14 Cortical alpha oscillations as a tool for auditory selective inhibition**
Antje Strauß, Malte Wöstmann and Jonas Obleser
- 21 Induction of plasticity in the human motor cortex by pairing an auditory stimulus with TMS**
Paul F. Sowman, Søren S. Dueholm, Jesper H. Rasmussen and Natalie Mrachacz-Kersting
- 27 Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: A hypothesis and theory**
Andrew C. Etchell, Blake W. Johnson and Paul F. Sowman
- 37 Sequencing at the syllabic and supra-syllabic levels during speech perception: An fMRI study**
Isabelle Deschamps and Pascale Tremblay
- 51 Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences**
Sylvia Vitello, Jane E. Warren, Joseph T. Devlin and Jennifer M. Rodd
- 65 Listening effort and accented speech**
Kristin J. Van Engen and Jonathan E. Peelle
- 69 Relationship between perceptual learning in speech and statistical learning in younger and older adults**
Thordis M. Neger, Toni Rietveld and Esther Janse
- 86 Effects of navigated TMS on object and action naming**
Julio C. Hernandez-Pavon, Niko Mäkelä, Henri Lehtinen, Pantelis Lioumis and Jyrki P. Mäkelä
- 95 Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity**
Keke Yu, Ruiming Wang, Li Li and Ping Li
- 104 Anterior insular cortex activity to emotional salience of voices in a passive oddball paradigm**
Chenyi Chen, Yu-Hsuan Lee and Yawei Cheng
- 115 Talking hands: Tongue motor excitability during observation of hand gestures associated with words**
Naeem Komeilipoor, Carmelo Mario Vicario, Andreas Daffertshofer and Paola Cesari
- 122 The neural processing of foreign-accented speech and its relationship to listener bias**
Han-Gyol Yi, Rajka Smiljanic and Bharath Chandrasekaran
- 134 Multi-talker background and semantic priming effect**
Marie Dekerle, Véronique Boulenger, Michel Hoen and Fanny Meunier

Editorial: Current research and emerging directions on the cognitive and neural organization of speech processing

Patti Adank^{1*}, Carolyn McGettigan² and Sonja A. E. Kotz^{3,4}

¹ Division of Psychology and Language Sciences, Speech, Hearing and Phonetic Sciences, University College London, London, UK, ² Department of Psychology, Royal Holloway University of London, Egham, UK, ³ Max Planck Institute Leipzig, Leipzig, Germany, ⁴ School of Psychological Sciences, University of Manchester, Manchester, UK

Keywords: speech perception, speech production, functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), electroencephalography, transcranial magnetic stimulation (TMS)

This Research Topic consists of 14 manuscripts discussing the cognitive and neural organization of speech processing. The contributions are grouped around four themes: (1) Spoken language comprehension under difficult listening conditions; (2) Sub-lexical processing; (3) Sensorimotor processing of speech; (4) Speech production.

Seven papers addressed speech perception under challenging listening conditions. Van Engen and Peelle (2014) discuss the effects of processing speech in an unfamiliar regional or foreign accent. They argue that, as perceiving accented speech incurs a processing cost, just like other types of distortions such as background noise, it should also be regarded as representing a challenging listening condition. Neger et al. (2014) focused on plasticity of speech processing in statistical and perceptual learning tasks in aging. They conclude that perceptual and statistical learning share mechanisms of implicit regularity detection, but that the ability to detect statistical regularities is impaired in older adults for fast visual sequences. Dekerle et al. (2014) examined whether speech perception in a multi-speaker background relies on semantic interference between the background and target speaker using a semantic priming paradigm in three experiments. Their results indicate that higher-level linguistic processes such as semantic priming may not be as automatic as commonly thought but are subjected to the limits of cognitive resources such as working memory and attention. Yi et al. (2014) evaluate how processing of foreign-accented speech relates to social cognition. It was concluded that foreign-accented speech perception engages greater activation of neural systems underlying speech perception, and that implicit Asian-foreign association is related to with decreased neural efficiency in early spectrotemporal processing. Vitello et al. (2014) used fMRI to address the question of how semantic ambiguities are resolved during speech comprehension.

Strauß et al. (2014) examined through literature review whether neural oscillations in the alpha frequency range (~10 Hz) act as a neural mechanism to selectively inhibit the processing of noise to improve auditory selective attention to task-relevant speech signals. Ding and Simon (2014) discuss whether cortical entrained activity is related more closely to speech perception or to auditory encoding that is not specific to speech, by reviewing evidence regarding various hypotheses about the functional roles of cortical entrainment to speech.

Three papers focused on perception of speech at sub-lexical levels. Deschamps and Tremblay (2014) studied perception of sub-lexical information by examining the neural bases of processing of simple syllables and more complex syllabic structures using fMRI, while Yu et al. (2014) used MEG to study the neural processing of disgust in anterior insula by presenting listeners with syllables with differed intended emotional meanings. Finally Chen et al. (2014) investigated processing of

OPEN ACCESS

Edited and reviewed by:

Hauke R. Heekeren,
Freie Universität Berlin, Germany

*Correspondence:

Patti Adank,
p.adank@ucl.ac.uk

Received: 16 April 2015

Accepted: 12 May 2015

Published: 27 May 2015

Citation:

Adank P, McGettigan C and Kotz SAE
(2015) Editorial: Current research and
emerging directions on the cognitive
and neural organization of speech
processing.
Front. Hum. Neurosci. 9:305.
doi: 10.3389/fnhum.2015.00305

acoustic and phonological information in lexical tones in Mandarin Chinese using EEG.

Two papers addressed sensorimotor processing of speech. Komeilipoor et al. (2014) report higher motor excitability as measured using Transcranial Magnetic Stimulation (TMS) in the tongue area during the presentation of meaningful gestures (noun-associated). Sowman et al. (2014) demonstrate that appropriately timed TMS to the hand area, paired with auditorily mediated excitation of the motor cortex, induces an enhancement of motor cortex excitability that lasts beyond the time of stimulation.

Two papers focused on speech production. Etchell et al. (2014) provide a review of the stuttering literature and Hernandez-Pavon et al. (2014) present a neuronavigated TMS study

exploring the neural locus of aspects of picture naming in healthy participants.

This Frontiers Research Topic allows new insights into the neurobiology of speech perception and production, and demonstrates how the field of speech science is now addressing issues at its very core. We believe that the future of the research in the field lies in the effective combination of research methods, e.g., EEG and TMS, or fMRI and EEG, as research will benefit from the strengths of each method. In conclusion, this Research Topic consists of 14 excellent contributions, and we are convinced the Topic will provide readers with novel ideas for future studies that will elucidate the cognitive and neural architecture of speech processing.

References

- Chen, C., Lee, Y., and Cheng, Y. (2014). Anterior insular cortex activity to emotional salience of voices in a passive oddball paradigm. *Front. Hum. Neurosci.* 8:743. doi: 10.3389/fnhum.2014.00743
- Dekerle, M., Boulenger, V., Hoen, M., and Meunier, F. (2014). Multi-talker background and semantic priming effect. *Front. Hum. Neurosci.* 8:878. doi: 10.3389/fnhum.2014.00878
- Deschamps, I., and Tremblay, F. (2014). Sequencing at the syllabic and supra-syllabic levels during speech perception: an fMRI study. *Front. Hum. Neurosci.* 8:492. doi: 10.3389/fnhum.2014.00492
- Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311
- Etchell, A. C., Johnson, B. W., and Sowman, P. F. (2014). Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: A hypothesis and theory. *Front. Hum. Neurosci.* 8:467. doi: 10.3389/fnhum.2014.00467
- Hernandez-Pavon, J. C., Mäkelä, N., Lehtinen, H., Lioumis, P., and Mäkelä, J. (2014). Effects of navigated TMS on object and action naming. *Front. Hum. Neurosci.* 8:660. doi: 10.3389/fnhum.2014.00660
- Komeilipoor, N., Vicario, C. M., Daffertshofer, A., and Cerari, P. (2014). Talking hands: tongue motor excitability during observation of hand gestures associated with words. *Front. Hum. Neurosci.* 8:767. doi: 10.3389/fnhum.2014.00767
- Neger, T. M., Rietveld, T., and Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Front. Hum. Neurosci.* 8:628. doi: 10.3389/fnhum.2014.00628
- Sowman, P. F., Rasmussen, J., Dueholm, S., and Mrachacz-Kersting, N. (2014). Induction of plasticity in the human motor cortex by pairing an auditory stimulus with TMS. *Front. Hum. Neurosci.* 8:398. doi: 10.3389/fnhum.2014.00398
- Strauß, A., Wöstmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8:350. doi: 10.3389/fnhum.2014.00350
- Van Engen, K. J., and Peelle, J. E. (2014). Listening effort and accented speech. *Front. Hum. Neurosci.* 8:577. doi: 10.3389/fnhum.2014.00577
- Vitello, S., Warren, J. E., Devlin, J. T., and Rodd, J. M. (2014). Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences. *Front. Hum. Neurosci.* 8:530. doi: 10.3389/fnhum.2014.00530
- Yi, H., Smiljanic, R., and Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Front. Hum. Neurosci.* 8:768. doi: 10.3389/fnhum.2014.00768
- Yu, K., Wang, R., Li, L., and Li, P. (2014). Processing of acoustic and phonological information of lexical tones in mandarin chinese revealed by mismatch negativity. *Front. Hum. Neurosci.* 8:729. doi: 10.3389/fnhum.2014.00729

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Adank, McGettigan and Kotz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cortical entrainment to continuous speech: functional roles and interpretations

Nai Ding^{1*} and Jonathan Z. Simon^{2,3,4*}

¹ Department of Psychology, New York University, New York, NY, USA

² Department of Electrical and Computer Engineering, University of Maryland College Park, College Park, MD, USA

³ Department of Biology, University of Maryland College Park, College Park, MD, USA

⁴ Institute for Systems Research, University of Maryland College Park, College Park, MD, USA

Edited by:

Sonja A. E. Kotz, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Reviewed by:

István Winkler, University of Szeged, Hungary

Jonas Obleser, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

*Correspondence:

Nai Ding, Department of Psychology, New York University, New York, NY 10012, USA

e-mail: gahding@gmail.com;

Jonathan Z. Simon, Department of Electrical and Computer Engineering, University of Maryland College Park, College Park, MD 20742, USA
e-mail: jzsimon@umd.edu

Auditory cortical activity is entrained to the temporal envelope of speech, which corresponds to the syllabic rhythm of speech. Such entrained cortical activity can be measured from subjects naturally listening to sentences or spoken passages, providing a reliable neural marker of online speech processing. A central question still remains to be answered about whether cortical entrained activity is more closely related to speech perception or non-speech-specific auditory encoding. Here, we review a few hypotheses about the functional roles of cortical entrainment to speech, e.g., encoding acoustic features, parsing syllabic boundaries, and selecting sensory information in complex listening environments. It is likely that speech entrainment is not a homogeneous response and these hypotheses apply separately for speech entrainment generated from different neural sources. The relationship between entrained activity and speech intelligibility is also discussed. A tentative conclusion is that theta-band entrainment (4–8 Hz) encodes speech features critical for intelligibility while delta-band entrainment (1–4 Hz) is related to the perceived, non-speech-specific acoustic rhythm. To further understand the functional properties of speech entrainment, a splitter's approach will be needed to investigate (1) not just the temporal envelope but what specific acoustic features are encoded and (2) not just speech intelligibility but what specific psycholinguistic processes are encoded by entrained cortical activity. Similarly, the anatomical and spectro-temporal details of entrained activity need to be taken into account when investigating its functional properties.

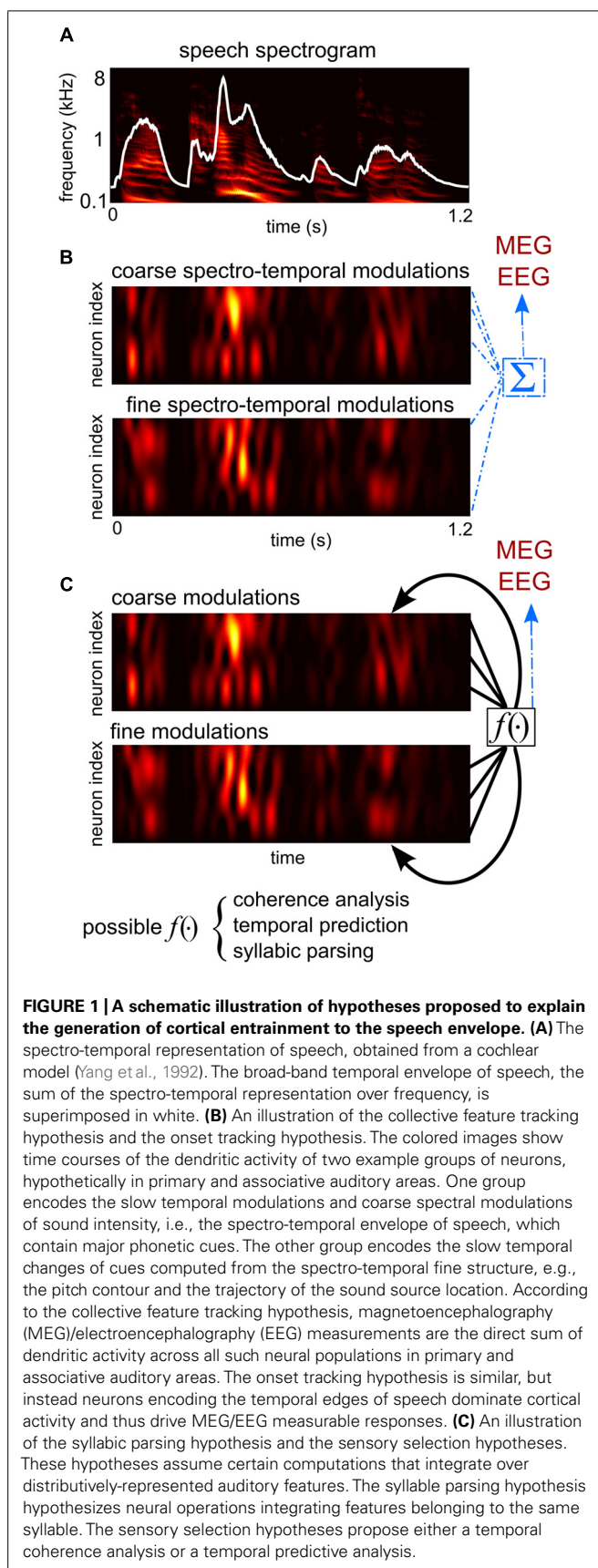
Keywords: auditory cortex, entrainment of rhythms, speech intelligibility, speech perception in noise, speech envelope, cocktail party problem

INTRODUCTION

Speech recognition is a process that maps an acoustic signal onto the underlying linguistic meaning. The acoustic properties of speech are complex and contain temporal dynamics on several time scales (Rosen, 1992; Chi et al., 2005). The time scale most critical for speech recognition is on the order of hundreds of milliseconds (1–10 Hz), and the temporal fluctuations on this time scale are usually called the *temporal envelope* (Figure 1A). Single neuron neurophysiology from animal models has shown that neurons in primary auditory cortex encode the analogous temporal envelope of other non-speech sounds by phase locked neural firing (Wang et al., 2003). In contrast, the finer scale acoustic properties that decide the pitch and timbre of speech at each time moment (acoustic fragments lasting a few 100 ms) are likely to be encoded using a spatial code, by either individual neurons (Bendor and Wang, 2005) or spatial patterns of cortical activity (Walker et al., 2011).

In the last decade or so, cortical entrainment to the temporal envelope of speech has been demonstrated in humans using magnetoencephalography (MEG; Ahissar et al., 2001; Luo and Poeppel, 2007), electroencephalography (EEG; Aiken and Picton, 2008), and electrocorticography (ECoG; Nourski et al., 2009). This envelope following response can be recorded from subjects

listening to sentences or spoken passages and therefore provides an online marker of neural processing of continuous speech. Envelope entrainment has mainly been seen in the waveform of low-frequency neural activity (<8 Hz) and in the power envelope of high-gamma activity (Pasley et al., 2012; Zion Golumbic et al., 2013). Although the phenomenon of envelope entrainment has been well established, its underlying neural mechanisms, and functional roles remain controversial. It is still under debate whether entrained cortical activity is more closely tied to the physical properties of the acoustic stimulus or to higher level language related processing that is directly related to speech perception. A number of studies have shown that cortical entrainment to speech is strongly modulated by top-down cognitive functions such as attention (Kerlin et al., 2010; Ding and Simon, 2012a; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013) and therefore is not purely a bottom-up response. On the other hand, cortical entrainment to the sound envelope is seen for non-speech sound (Lalor et al., 2009; Hämäläinen et al., 2012; Millman et al., 2012; Wang et al., 2012; Steinschneider et al., 2013) and therefore does not rely on speech-specific neural processing. In this article, we first summarize a number of hypotheses about the functional roles of envelope entrainment, and then review the literature about how envelope entrainment is affected by speech intelligibility.



FUNCTIONAL ROLES OF CORTICAL ENTRAINMENT

A number of hypotheses have been proposed about what aspects of speech, ranging from its acoustic features to its linguistic meaning, are encoded by entrained cortical activity. A few dominant hypotheses are summarized and compared (**Table 1**). Other unresolved questions about cortical neural entrainment, e.g., what the biophysical mechanisms generating cortical entrainment are, and whether entrained neural activity is related to spontaneous neural oscillations, are not covered here (see discussions in e.g., Schroeder and Lakatos, 2009; Howard and Poeppel, 2012; Ding and Simon, 2013b).

ONSET TRACKING HYPOTHESIS

Speech is dynamic and is full of acoustic “edges,” e.g., onsets and offsets. These edges usually occur at syllable boundaries and are well characterized by the speech envelope. It is well known that a reliable macroscopic brain response can be evoked by an acoustic edge. Therefore, it has been proposed that neural entrainment to the speech envelope is a superposition of discrete, edge/onset related brain responses (Howard and Poeppel, 2010). Consistent with this hypothesis, it has been shown that the sharpness of acoustic edges, i.e., how quickly sound intensity increases, strongly influences cortical tracking of the sound envelope (Prendergast et al., 2010; Doelling et al., 2014). A challenge of this hypothesis, however, is that speech is continuously changing and it remains a problem as to which acoustic transients can be counted as edges.

If this hypothesis is true, a question naturally follows about whether envelope entrainment can provide insights that cannot be learned using the traditional event-related response approach. The answer is yes. Cortical responses, including edge/onset related auditory evoked responses, are stimulus-dependent, and quickly adapt to the spectro-temporal structure of the stimulus (Zacharias et al., 2012; Herrmann et al., 2014). Therefore, even if envelope entrainment is just a superposition of event-related responses, it can still provide insights about the properties of cortical activity when it is adapted to the acoustic properties of speech.

COLLECTIVE FEATURE TRACKING HYPOTHESIS

When sound enters the ear, it is decomposed into narrow frequency bands in the auditory periphery and is further decomposed into multi-scale acoustic features in the central auditory system, such as pitch, sound source location information, and coarse spectro-temporal modulations (Shamma, 2001; Ghitza et al., 2012). In speech, most acoustic features coherently fluctuate in time and these coherent fluctuations are captured by the speech envelope. If a neuron or a neural population encodes an acoustic feature, its activity is synchronized to the strength of that acoustic feature. As a result, neurons or neural networks that are tuned to coherently fluctuating speech features are activated coherently (Shamma et al., 2011).

Analogously to the speech envelope being the summation of the power of all speech features at each time moment, the large-scale neural entrainment to speech measured by MEG/EEG can be the summation of neural activity tracking different acoustic features of speech (**Figure 1B**). It is therefore plausible to hypothesize that macroscopic speech entrainment is a passive summation of microscopic neural tracking of acoustic features

Table 1 | A summary of major hypotheses about the functional roles of cortical entrainment to speech.

Hypothesis	Underlying neural computations	Reference
Onset tracking	Temporal edge detection	Howard and Poeppel (2010)
Collective feature tracking	Spectro-temporal feature coding	Ding and Simon (2012b), Ghitza et al. (2012)
Syllabic parsing	Binding features of the same syllable; discretization	Giraud and Poeppel (2012), Ghitza (2013)
Sensory selection I	Temporal coherence-based binding of auditory features	Shamma et al. (2011), Ding and Simon (2012a)
Sensory selection II	Modulation of neuronal excitability; temporal prediction	Schroeder and Lakatos (2009)

across neurons/networks (Ding and Simon, 2012b). Based on this hypothesis, the MEG/EEG speech entrainment is a marker of a collective cortical representation of speech but does not play any additional roles in regulating neuronal activity.

The onset tracking hypothesis can be viewed as a special case of the collective feature tracking hypothesis, when the acoustic features driving cortical responses are restricted to a set of discrete edges. The collective feature tracking hypothesis, however, is more general since it allows features to be continuously changing and also incorporates features that are not associated with sharp intensity changes, such as changes in the pitch contour (Oblaser et al., 2012), and sound source location. Under the onset tracking hypothesis, entrained neural activity is a superposition of onset/edge-related auditory evoked responses. Under the more general collective feature tracking hypothesis, at a first-order approximation, entrained activity is a convolution between speech features, e.g., the temporal envelopes in different narrow frequency bands, and the corresponding response functions, e.g., the response evoked by a very brief tone pip in the corresponding frequency band (Lalor et al., 2009; Ding and Simon, 2012b).

SYLLABIC PARSING HYPOTHESIS

During speech recognition, the listener must segment a continuous acoustic signal into a sequence of discrete linguistic symbols, into the units of, e.g., phonemes, syllables or words. The boundaries between phonemes, and especially syllables, are relatively well encoded by the speech envelope (Stevens, 2002; Ghitza, 2013, see also Cummins, 2012). Furthermore, the average syllabic rate ranges between 5 and 8 Hz across languages (Pellegrino et al., 2011) and the rate for stressed syllables is below 4 Hz for English (Greenberg et al., 2003). Therefore it has been hypothesized that neural entrainment to the speech envelope plays a role in creating a syllabic level, discrete, representation of speech (Giraud and Poeppel, 2012). In particular, it has been hypothesized that each cycle of the cortical theta oscillation (4–8 Hz) is aligned to the portion of speech signal in between of two vowels, corresponding to two adjacent peaks in the speech envelope. Auditory features within a cycle of theta oscillation are then used to decode the phonetic information of speech (Ghitza, 2011, 2013). Therefore, according to this hypothesis, speech entrainment does not only passively track acoustic features but also reflects the language-based packaging of speech into syllable size chunks. Since syllables play different roles in segmenting syllable-timed language and stress-timed language (Cutler et al., 1986), further cross-language research may

further elucidate which of these neural processes are represented in envelope tracking activity.

SENSORY SELECTION HYPOTHESIS

In everyday listening environments, speech is often embedded in a complex acoustic background. Therefore, to understand speech, a listener must segregate speech from the listening background and process it selectively. A useful strategy for the brain would be to find and selectively process moments in time (or spectro-temporal instances in a more general framework) that are dominated by speech and ignore the moments dominated by the background (Wang, 2005; Cooke, 2006). In other words, the brain might robustly encode speech by taking glimpses at the temporal (or spectro-temporal) features that contain critical speech information. The rhythmicity of speech (Schroeder and Lakatos, 2009; Giraud and Poeppel, 2012), and the temporal coherence between acoustic features (Shamma et al., 2011), are both reflected by the speech envelope and so become critical cues for the brain to decide where the useful speech information lies. Therefore, envelope entrainment may play a critical role in the neural segregation of speech and the listening background.

In a complex listening environment, cortical entrainment to speech has been found to be largely invariant to the listening background (Ding and Simon, 2012a; Ding and Simon, 2013a). Two possible functional roles have been hypothesized for the observed background-invariant envelope entrainment. One is that the brain uses temporal coherence to bind together acoustic features belonging to the same speech stream and envelope entrainment may reflect computations related to this coherence analysis (Shamma et al., 2011; Ding and Simon, 2012a). The other is that envelope entrainment is used by the brain to predict which moments contain more information about speech than the acoustic background and then guide the brain to selectively process those moments (Schroeder et al., 2008; Schroeder and Lakatos, 2009; Zion Golumbic et al., 2012).

WHICH HYPOTHESIS IS TRUE? AN ANALYSIS-BY-SYNTHESIS ACCOUNT OF SPEECH PROCESSING

Speech processing is a complicated process that can be roughly divided into an analysis stage and a synthesis stage. In the analysis stage, speech sounds are decomposed into primitive auditory features, a process that starts from the cochlea and applies mostly equally to the auditory encoding of both speech and non-speech sounds. A later synthesis stage, in contrast, combines multiple

auditory features to create speech perception, including, e.g., binding spectro-temporal cues to determine phonemic categories, or integrating multiple acoustic cues to segregate a target speech stream from an acoustic background. The onset tracking hypothesis and the collective feature tracking hypothesis both view speech entrainment as a passive auditory encoding mechanism belonging to the analysis stage. Note, however, that the analysis stage does include some integration over separately represented features also. For example, neural processing of pitch and spectral modulations requires integrating information across frequency. Functionally, however, the purpose of integrating features in the analysis stage is to extract higher level auditory features rather than to construct linguistic/perceptual entities.

The syllabic parsing hypothesis and the sensory selection hypothesis propose functional roles of cortical entrainment in the synthesis stage. They hypothesize that cortical entrainment is involved in combining features into linguistic units, e.g., syllables, or perceptual units, e.g., speech streams (**Figure 1C**). These additional functional roles may be implemented in two ways: an active mechanism would be one that entrained cortical activity, as a large-scale voltage fluctuation, directly regulating syllabic parsing or sensory selection (Schroeder et al., 2008; Schroeder and Lakatos, 2009). A passive mechanism would be one where neural computations related to syllabic parsing or sensory selection would generate spatially coherent neural signals that are measurable by macroscopic recording tools.

Although clearly distinctive from each other, the four hypotheses may all be true for different functional areas of the brain and describe different neural generators for speech entrainment. Onset detection, feature tracking, syllabic parsing, and sensory selection are all neural computations necessary for speech recognition and all of them are likely to be synchronized to the speech rhythm carried by the envelope. Therefore, these neural computations may all be reflected by cortical entrainment to speech, and may only differ in their fine-scale neural generators. It remains unclear, however, whether these fine-scale neural generators can be resolved by macroscopic recording tools such as MEG and EEG.

Future studies are needed to explicitly test these hypotheses, or explicitly modify them, to determine which specific acoustic features and which specific psycholinguistic processes are relevant to cortical entrainment. For example, to dissociate the onset tracking hypothesis and the collective feature tracking hypothesis, one approach is to create explicit computational models for them and test which model would fit the data better. To test the syllabic parsing hypothesis, it will be important to calculate the correlation between cortical entrainment and relevant behavioral measures, e.g., misallocation of syllable boundaries (Woodfield and Akeroyd, 2010). To test the sensory selection hypothesis, stimuli that vary in their temporal probability or coherence among spectro-temporal features are likely to be revealing.

ENVELOPE ENTRAINMENT AND SPEECH INTELLIGIBILITY ENTRAINMENT AND ACOUSTIC MANIPULATION OF SPEECH

As indicated by its name, envelope entrainment is correlated with the speech envelope, an acoustic property of speech. Nevertheless, neural encoding of speech must underlie the ultimate goal of

decoding its meaning. Therefore, it is critical to identify if cortical entrainment to speech is related to any behavioral measure during speech recognition, such as speech intelligibility.

A number of studies have compared cortical activity entrained to intelligible speech and unintelligible speech. One approach is to vary the acoustic stimulus and analyze how cortical entrainment changes within individual subjects. Some studies have found that cortical entrainment to normal sentences is similar to cortical entrainment to sentences that are played backward in time (Howard and Poeppel, 2010; Peña and Melloni, 2012; though see Gross et al., 2013).

A second way to reduce intelligibility is to introduce different types of acoustic interference. When speech is presented together with stationary noise, delta-band (1–4 Hz) cortical entrainment to the speech is found to be robust to noise until the listeners can barely hear speech, while theta-band (4–8 Hz) entrainment decreases gradually as the noise level increases (Ding and Simon, 2013a). In this way, theta-band entrainment is correlated with noise level and also speech intelligibility, but delta-band entrainment is not. When speech is presented together with a competing speech stream, cortical entrainment is found to be robust against the level of the competing speech stream even though intelligibility drops (Ding and Simon, 2012a; theta- and delta-band activity was not analyzed separately there).

A third way to reduce speech intelligibility is to degrade the spectral resolution through noise-vocoding, which destroys spectro-temporal fine structure but preserves the temporal envelope (Shannon et al., 1995). When the spectral resolution of speech decreases, it has been shown that theta-band cortical entrainment reduces (Peelle et al., 2013; Ding et al., 2014) but delta-band entrainment enhances (Ding et al., 2014). In contrast, when background noise is added to speech and the speech-noise mixture is noise vocoded, it is found that both delta- and theta-band entrainment is reduced by vocoding (Ding et al., 2014).

A fourth way to vary speech intelligibility is to directly manipulate the temporal envelope (Doelling et al., 2014). When the temporal envelope in the delta-theta frequency range is corrupted, cortical entrainment in the corresponding frequency bands degrades and so does speech intelligibility. When a processed speech envelope is used to modulate a broadband noise carrier, the stimulus is not intelligible but reliable cortical entrainment is nevertheless seen.

In many of these studies investigating the correlation between cortical entrainment and intelligibility, a common issue is that stimuli which differ in intelligibility also differ in acoustic properties. This makes it difficult to determine if changes in cortical entrainment arise from changes in speech intelligibility or from changes in acoustic properties. For example, speech syllables generally have a sharper onset than offset, so reversing speech in time changes those temporal characteristics. Similarly, when the spectral resolution is reduced, neurons tuned to fine spectral features are likely to be deactivated. Therefore, based on the studies reviewed here, it can only be tentatively concluded that, when critical speech features are manipulated, speech intelligibility, and theta-band entrainment are affected in similar ways while delta-band entrainment is not. It remains unclear about

whether speech intelligibility causally modulates cortical entrainment or that auditory encoding, reflected by cortical entrainment, influences downstream language processing and therefore become indirectly related to intelligibility.

VARIABILITY BETWEEN LISTENERS

A second approach to address the correlation between neural entrainment and speech intelligibility is to investigate the variability across listeners. Peña and Melloni (2012) compared neural responses in listeners who speak the tested language and listeners who do not speak the tested language. It was found that language understanding does not significantly change the low-frequency neural responses, but it does change high-gamma band neural activity. Within the group of native speakers, the intelligibility score still varied broadly in the challenging listening conditions. Delta-band, but not theta-band, cortical entrainment has been shown to correlate with intelligibility scores for individual listeners in a number of studies (Ding and Simon, 2013a; Ding et al., 2014; Doelling et al., 2014). The advantage of investigating inter-subject variability is that it avoids modifications of the sound stimuli. Nevertheless, it still cannot identify whether the individual differences in speech recognition arise from the individual differences in auditory processing (Ruggles et al., 2011), language related processing, or cognitive control.

The speech intelligibility approach in general, suffers from a drawback that it is the end point of the entire speech recognition chain, and is not targeted at specific linguistic computations, e.g., allocating the boundaries between syllables. Furthermore, when the acoustic properties of speech are degraded, speech recognition requires additional cognitive control and the involved neural processing networks adapt (Du et al., 2011; Wild et al., 2012; Erb et al., 2013; Lee et al., 2014). Therefore, just from a change in speech intelligibility, it is difficult to trace what kinds of neural processing are affected.

DISTINCTIONS BETWEEN DELTA- AND THETA-BAND ENTRAINMENT

In summary of these different approaches, when the acoustic properties of speech are manipulated, theta-band entrainment often shows changes that correlate with speech intelligibility. For the same stimulus, however, the speech intelligibility measured from individual listeners is often correlated with delta-band entrainment. To explain this dichotomy, here we hypothesize that theta-band entrainment encodes syllabic-level acoustic features critical for speech recognition, while delta-band entrainment is more closely related to the perceived acoustic rhythm rather than the phonemic information of speech. This hypothesis is also consistent with the fact that speech modulations between 4 and 8 Hz are critical for intelligibility (Drullman et al., 1994a,b; Elliott and Theunissen, 2009) while temporal modulations below 4 Hz include prosodic information of speech (Goswami and Leong, 2013) and it is the frequency range important for music rhythm perception (Patel, 2008; Farbood et al., 2013).

ENVELOPE ENTRAINMENT TO NON-SPEECH SOUNDS

Although speech envelope entrainment may show correlated changes with speech intelligibility when the acoustic properties

of speech are manipulated, speech intelligibility is probably not a major driving force for envelope entrainment. A critical evidence is that envelope entrainment can be observed for non-speech sounds in humans and both speech and non-speech sounds in animals. Here, we briefly review human studies on envelope entrainment for non-speech sounds (see e.g., Steinschneider et al., 2013 for a comparison between envelope entrainment in human and animal models).

Traditionally, envelope entrainment has been studied using the auditory steady-state response (aSSR), a periodic neural response tracking the stimulus repetition rate or modulation rate. An aSSR at a given frequency can be elicited by, e.g., a click or tone-pip train repeating at the same frequency (Nourski et al., 2009; Xiang et al., 2010), and by amplitude or frequency modulation at that frequency (Picton et al., 1987; Ross et al., 2000; Wang et al., 2012). Although the cortical aSSR can be elicited in a broad frequency range (up to ~100 Hz), speech envelope entrainment is likely to be related to the slow aSSR in the corresponding frequency range, i.e., below 10 Hz (see Picton, 2007 for a review of the robust aSSR of 40 Hz and above). More recently, cortical entrainment has also been demonstrated for sounds modulated by an irregular envelope (Lalor et al., 2009). Low-frequency (<10 Hz) cortical entrainment to non-speech sound shares many properties with cortical entrainment to speech. For example, when envelope entrainment is modeled using a linear system-theoretic model, the neural response is qualitatively similar for speech (Power et al., 2012) and amplitude-modulated tones (Lalor et al., 2009). Furthermore, low-frequency (<10 Hz) cortical entrainment to non-speech sound is also strongly modulated by attention (Elhilali et al., 2009; Power et al., 2010; Xiang et al., 2010), and the phase of entrained activity is predictive of listeners' performance in some sound-feature detection tasks (Henry and Obleser, 2012; Ng et al., 2012).

SUMMARY

Cortical entrainment to the speech envelope provides a powerful tool to investigate online neural processing of continuous speech. It greatly extends the traditional event-related approach that can only be applied to analyze the response to isolated syllables or words. Although envelope entrainment has attracted researchers' attention in the last decade, it is still a less well-characterized cortical response than event-related responses. The basic phenomenon of envelope entrainment has been reliably seen in EEG, MEG, and ECoG, even at the single-trial level (Ding and Simon, 2012a; O'Sullivan et al., 2014). Hypotheses have been proposed about the neural mechanisms generating cortical entrainment and its functional roles, but these hypotheses remain to be explicitly tested. To test these hypotheses, a computational modeling approach is likely to be effective. For example, rather than just calculating the correlation between neural activity and the speech envelope, more explicit computational models can be proposed and used to fit the data (e.g., Ding and Simon, 2013a). Furthermore, to understand what linguistic computations are achieved by entrained cortical activity, more fine-scaled behavioral measures are likely to be required, e.g., measures related to syllable boundary allocation rather than the general measure of intelligibility. Finally, the

anatomical, temporal, and spectral specifics of cortical entrainment should be taken into account when discussing its functional roles (Peña and Melloni, 2012; Zion Golumbic et al., 2013; Ding et al., 2014).

AUTHOR CONTRIBUTIONS

Nai Ding and Jonathan Z. Simon wrote and approved the paper.

ACKNOWLEDGMENT

The work is supported by NIH grant R01 DC 008342.

REFERENCES

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13367–13372. doi: 10.1073/pnas.201400998
- Aiken, S. J., and Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear Hear.* 29, 139–157. doi: 10.1097/AUD.0b013e31816453dc
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165. doi: 10.1038/nature03867
- Chi, T., Ru, P., and Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* 118, 887–906. doi: 10.1121/1.1945807
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.* 119, 1562–1573. doi: 10.1121/1.2166600
- Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Front. Psychol.* 3:364. doi: 10.3389/fpsyg.2012.00364
- Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *J. Mem. Lang.* 25, 385–400. doi: 10.1016/0749-596X(86)90033-1
- Ding, N., Chatterjee, M., and Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88C, 41–46. doi: 10.1016/j.neuroimage.2013.10.054 [Epub ahead of print].
- Ding, N., and Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109
- Ding, N., and Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi: 10.1152/jn.00297.2011
- Ding, N., and Simon, J. Z. (2013a). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi: 10.1523/JNEUROSCI.5297-12.2013
- Ding, N., and Simon, J. Z. (2013b). Power and phase properties of oscillatory neural responses in the presence of background activity. *J. Comput. Neurosci.* 34, 337–343. doi: 10.1007/s10827-012-0424-6
- Doelling, K., Arnal, L., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768. doi: 10.1016/j.neuroimage.2013.06.035
- Drullman, R., Festen, J. M., and Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* 95, 2670–2680. doi: 10.1121/1.409836
- Drullman, R., Festen, J. M., and Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064. doi: 10.1121/1.408467
- Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X., Li, L., et al. (2011). Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. *Cereb. Cortex* 21, 698–707. doi: 10.1093/cercor/bhq136
- Elhilali, M., Xiang, J., Shamma, S. A., and Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol.* 7:e1000129. doi: 10.1371/journal.pbio.1000129
- Elliott, T., and Theunissen, F. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5:e1000302. doi: 10.1371/journal.pcbi.1000302
- Erb, J., Henry, M. J., Eisner, F., and Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *J. Neurosci.* 33, 10688–10697. doi: 10.1523/JNEUROSCI.4596-12.2013
- Farbood, M. M., Marcus, G., and Poeppel, D. (2013). Temporal dynamics and the identification of musical key. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 911–918. doi: 10.1037/a0031087
- Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130
- Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4:138. doi: 10.3389/fpsyg.2013.00138
- Ghitza, O., Giraud, A.-L., and Poeppel, D. (2012). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Front. Hum. Neurosci.* 6:340. doi: 10.3389/fnhum.2012.00340
- Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Goswami, U., and Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives? *Lab. Phonol.* 4, 67–92. doi: 10.1515/lp-2013-0004
- Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech – a syllable-centric perspective. *J. Phon.* 31, 465–485. doi: 10.1016/j.wocn.2003.09.005
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752
- Hämäläinen, J. A., Rupp, A., Soltész, F., Szűcs, D., and Goswami, U. (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: an MEG study. *Neuroimage* 59, 2952–2961. doi: 10.1016/j.neuroimage.2011.09.075
- Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20095–20100. doi: 10.1073/pnas.1213390109
- Herrmann, B., Schlichting, N., and Obleser, J. (2014). Dynamic range adaptation to spectral stimulus statistics in human auditory cortex. *J. Neurosci.* 34, 327–331. doi: 10.1523/JNEUROSCI.3974-13.2014
- Howard, M. F., and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104, 2500–2511. doi: 10.1152/jn.00251.2010
- Howard, M. F., and Poeppel, D. (2012). The neuromagnetic response to spoken sentences: co-modulation of theta band amplitude and phase. *Neuroimage* 60, 2118–2127. doi: 10.1016/j.neuroimage.2012.02.028
- Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *J. Neurosci.* 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010
- Lalor, E. C., Power, A. J., Reilly, R. B., and Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.* 102, 349–359. doi: 10.1152/jn.09086.2008
- Lee, A. K., Larson, E., Maddox, R. K., and Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hear. Res.* 307, 111–120. doi: 10.1016/j.heares.2013.06.010
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020
- Millman, R. E., Prendergast, G., Hymers, M., and Green, G. G. (2012). Representations of the temporal envelope of sounds in human auditory cortex: can the results from invasive intracortical “depth” electrode recordings be replicated using non-invasive MEG “virtual electrodes”? *Neuroimage* 64, 185–196. doi: 10.1016/j.neuroimage.2012.09.017
- Ng, B. S. W., Schroeder, T., and Kayser, C. (2012). A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *J. Neurosci.* 32, 12268–12276. doi: 10.1523/JNEUROSCI.1877-12.2012
- Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564–15574. doi: 10.1523/JNEUROSCI.3065-09.2009

- Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Front. Hum. Neurosci.* 6:250. doi: 10.3389/fnhum.2012.00250
- O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex*. doi: 10.1093/cercor/bht355
- Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol.* 10:e1001251. doi: 10.1371/journal.pbio.1001251
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.
- Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118
- Pellegrino, F., Coupé, C., and Marsico, E. (2011). Across-language perspective on speech information rate. *Language* 87, 539–558. doi: 10.1353/lan.2011.0057
- Peña, M., and Melloni, L. (2012). Brain oscillations during spoken sentence processing. *J. Cogn. Neurosci.* 24, 1149–1164. doi: 10.1162/jocn_a_00144
- Picton, T. W. (2007). "Audiometry using auditory steady-state responses," in *Auditory Evoked Potentials: Basic Principles and Clinical Application*, eds. R. F. Burkard, J. J. Eggermont, and M. Don (Baltimore: Lippincott Williams & Wilkins), 441–462.
- Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J. C., and Maiste, A. C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J. Acoust. Soc. Am.* 82, 165–178. doi: 10.1121/1.395560
- Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., and Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *Eur. J. Neurosci.* 35, 1497–1503. doi: 10.1111/j.1460-9568.2012.08060.x
- Power, A. J., Lalor, E. C., and Reilly, R. B. (2010). Endogenous auditory spatial attention modulates obligatory sensory activity in auditory cortex. *Cereb. Cortex* 21, 1223–1230. doi: 10.1093/cercor/bhq233
- Prendergast, G., Johnson, S. R., and Green, G. G. (2010). Temporal dynamics of sinusoidal and non-sinusoidal amplitude modulation. *Eur. J. Neurosci.* 32, 1599–1607. doi: 10.1111/j.1460-9568.2010.07423.x
- Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070
- Ross, B., Borgmann, C., Draganova, R., Roberts, L. E., and Pantev, C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J. Acoust. Soc. Am.* 108, 679–691. doi: 10.1121/1.429600
- Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15516–15521. doi: 10.1073/pnas.1108912108
- Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Shamma, S. (2001). On the role of space and time in auditory processing. *Trends Cogn. Sci.* 5, 340–348. doi: 10.1016/S1364-6613(00)01704-6
- Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34, 114–123. doi: 10.1016/j.tins.2010.11.002
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303
- Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special? *Hear. Res.* 305, 57–73. doi: 10.1016/j.heares.2013.05.013
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* 111, 1872–1891. doi: 10.1121/1.1458026
- Walker, K. M., Bizley, J. K., King, A. J., and Schnupp, J. W. (2011). Multiplexed and robust representations of sound features in auditory cortex. *J. Neurosci.* 31, 14565–14576. doi: 10.1523/JNEUROSCI.2074-11.2011
- Wang, D. (2005). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation By Humans and Machines*, ed. P. Divenyi (New York: Springer), 181–197.
- Wang, X., Lu, T., and Liang, L. (2003). Cortical processing of temporal modulations. *Speech Commun.* 41, 107–121. doi: 10.1016/S0167-6393(02)00097-3
- Wang, Y., Ding, N., Ahmar, N., Xiang, J., Poeppel, D., and Simon, J. Z. (2012). Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. *J. Neurophysiol.* 107, 2033–2041. doi: 10.1152/jn.00310.2011
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012
- Woodfield, A., and Akeroyd, M. A. (2010). The role of segmentation difficulties in speech-in-speech understanding in older and hearing-impaired adults. *J. Acoust. Soc. Am.* 128, EL26–EL31. doi: 10.1121/1.3443570
- Xiang, J., Simon, J., and Elhilali, M. (2010). Competing streams at the cocktail Party: exploring the mechanisms of attention and temporal integration. *J. Neurosci.* 30, 12084–12093. doi: 10.1523/JNEUROSCI.0827-10.2010
- Yang, X., Wang, K., and Shamma, S. A. (1992). Auditory representations of acoustic signals. *IEEE Trans. Inf. Theory* 38, 824–839. doi: 10.1109/18.119739
- Zacharias, N., König, R., and Heil, P. (2012). Stimulation – history effects on the M100 revealed by its differential dependence on the stimulus onset interval. *Psychophysiology* 49, 909–919. doi: 10.1111/j.1469-8986.2012.01370.x
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., Mckhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037
- Zion Golumbic, E. M., Poeppel, D., and Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain Lang.* 122, 151–161. doi: 10.1016/j.bandl.2011.12.010

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 March 2014; accepted: 27 April 2014; published online: 28 May 2014.

Citation: Ding N and Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Ding and Simon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cortical alpha oscillations as a tool for auditory selective inhibition

Antje Strauß^{1*†}, Malte Wöstmann^{1,2†} and Jonas Obleser¹

¹ Max Planck Research Group "Auditory Cognition", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

² International Max Planck Research School on Neuroscience of Communication, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Edited by:

Carolyn McGettigan, Royal Holloway
University of London, UK

Reviewed by:

Johanna M. Zumer, University of
Birmingham, UK

Rebecca E. Millman, York
NeuroImaging Centre, UK

*Correspondence:

Antje Strauß, Max Planck Research
Group "Auditory Cognition", Max
Planck Institute for Human Cognitive
and Brain Sciences, Stephanstraße
1A, 04103 Leipzig, Germany
e-mail: strauss@cbs.mpg.de

[†] These authors have contributed
equally to this work.

Listening to speech is often demanding because of signal degradations and the presence of distracting sounds (i.e., "noise"). The question how the brain achieves the task of extracting only relevant information from the mixture of sounds reaching the ear (i.e., "cocktail party problem") is still open. In analogy to recent findings in vision, we propose cortical alpha (~10 Hz) oscillations measurable using M/EEG as a pivotal mechanism to selectively inhibit the processing of noise to improve auditory selective attention to task-relevant signals. We review initial evidence of enhanced alpha activity in selective listening tasks, suggesting a significant role of alpha-modulated noise suppression in speech. We discuss the importance of dissociating between noise interference in the auditory periphery (i.e., energetic masking) and noise interference with more central cognitive aspects of speech processing (i.e., informational masking). Finally, we point out the adverse effects of age-related hearing loss and/or cognitive decline on auditory selective inhibition. With this perspective article, we set the stage for future studies on the inhibitory role of alpha oscillations for speech processing in challenging listening situations.

Keywords: alpha, neural oscillations, effortful listening, inhibition, masking, speech, aging, hearing loss

1. INTRODUCTION

In ecological listening situations, auditory signals are rarely perceived in quiet due to the presence of different auditory maskers such as distracting background speech or environmental noise. Thus, sounds from different sources greatly overlap spectro-temporally at the level of the listener's ear. What are the neural correlates that facilitate selective listening to relevant target signals despite irrelevant auditory input (i.e., the "cocktail party problem"; Cherry, 1953)? At the central neural level, two complementary mechanisms of top-down control (i.e., regulation of subsidiary cognitive processes) should be considered: First, top-down selective attention to relevant information (Fritz et al., 2007) could facilitate target processing by enhancing the neural response to the attended stream (i.e., gain control; Lee et al., 2013). Second, top-down selective inhibition of maskers (Melara et al., 2002) could help to direct limited processing capacities away from irrelevant information (Desimone and Duncan, 1995), thereby avoiding full processing of distractors (Foxy and Snyder, 2011).

In this regard, interference of auditory maskers might be the result of both insufficient attention to the target and poor inhibition of noise and distractors. In this perspective article we focus on the latter, that is, neural mechanisms of auditory selective inhibition. We propose that cortical alpha (~10 Hz) oscillations are an important tool for top-down control as they regulate the inhibition of masker information during speech processing in challenging listening situations.

2. THE FUNCTIONAL SIGNIFICANCE OF ALPHA OSCILLATIONS

Neural oscillations in the alpha frequency range (~10 Hz) are the most dominant signal measurable in the human magneto- and electroencephalogram (M/EEG), going back to their first description by Berger (1931). The earliest observations of the alpha rhythm revealed that its amplitude is enhanced in humans who are awake but not actively engaged in any task. This finding led initially to the view that high alpha power might simply reflect the default state of brain inactivity or "cortical idling" (for a review, see Pfurtscheller et al., 1996).

Only within the last two decades, the functional significance of alpha oscillations has been recognized and furthermore its ubiquitous role across sensory modalities (visual: for review see Mathewson et al., 2011; sensorimotor: e.g., Haegens et al., 2012; auditory: e.g., Hartmann et al., 2012) and cognitive tasks (working memory: e.g., Jensen et al., 2002; attention: for a review see Klimesch, 2012; decision making: e.g., Cohen et al., 2009). One unifying mechanism suggested for alpha rhythms across modalities and brain areas is that it provides a neural means to functionally inhibit the processing of currently task-irrelevant or task-detrimental information (Jensen and Mazaheri, 2010; Foxy and Snyder, 2011). Please note that the opposite mechanism also has been proposed where higher inter-areal alpha phase synchronization does not index cortical inhibition but increased information processing such as for internal (working memory related) information processes (Palva and Palva, 2011). The functional inhibition hypothesis, though, has

received neurophysiological support. For example, both alpha power (i.e., squared amplitude) and alpha phase modulate neuronal spike rate (Haegens et al., 2011) and thus can directly affect the efficiency of neural information flow. In future, the alpha network needs to be further characterized by its phase-amplitude coupling to gamma oscillations (Jensen et al., 2012) and its role in top-down control as implemented in different cortical layers (Buffalo et al., 2011; Spaak et al., 2012) or in thalamo-cortical communication (Strauss et al., 2010; Roux et al., 2013).

Despite the abundance of studies on the role of alpha activity for visual selective inhibition, there are currently few studies that directly examine the role of alpha activity in the auditory modality. Recently, a series of studies found modulations in alpha power in a variety of auditory tasks prompted by degraded spectral detail (Obleser and Weisz, 2012), missing temporal expectations (Wilsch et al., 2014), working memory load (Leiberg et al., 2006; Obleser et al., 2012), or syntactic complexity (Meyer et al., 2013). Together, these findings provide good evidence that alpha oscillatory power can be a reliable indicator of auditory cognitive load (see also Luo et al., 2005; Kaiser et al., 2007). In the following section, we argue that part of this cognitive load occurs due to auditory selective inhibition as a compensatory mechanism for demanding listening situations and manifests in enhanced alpha power.

3. ALPHA OSCILLATIONS AS A TOOL FOR AUDITORY SELECTIVE INHIBITION

A common observation from our laboratory is a prominent increase in alpha power when participants listen to auditory materials presented against background noise (e.g., Wilsch et al., 2014). **Figure 1A**, for example, shows the grand average alpha power of 11 participants during a lexical decision task on isolated words presented in quiet (published in Strauss et al., 2014) and in white noise. For words in quiet, alpha power at around 10 Hz did not considerably increase after word onset. However, when words were presented in noise, alpha power was increased during the first 500 ms after word onset corresponding to the first two thirds of the average word duration. This effect was strongest over temporal and occipital sites (topography in **Figure 1A**) suggesting the inhibition of the task irrelevant visual modality but also compensatory mechanisms within speech-related areas. Critically, alpha power difference did not depend on ITPC (inter-trial phase coherence) differences, as indicated by the absence of a stronger ITPC in noise compared to quiet (**Figure 1B**). In fact, no significant ITPC differences were observed between 0.2 and 0.5 s. We therefore presume that induced (i.e., not strictly stimulus-locked; Freunberger et al., 2009) alpha power is crucial for speech processing in challenging listening conditions as it suppresses irrelevant information.

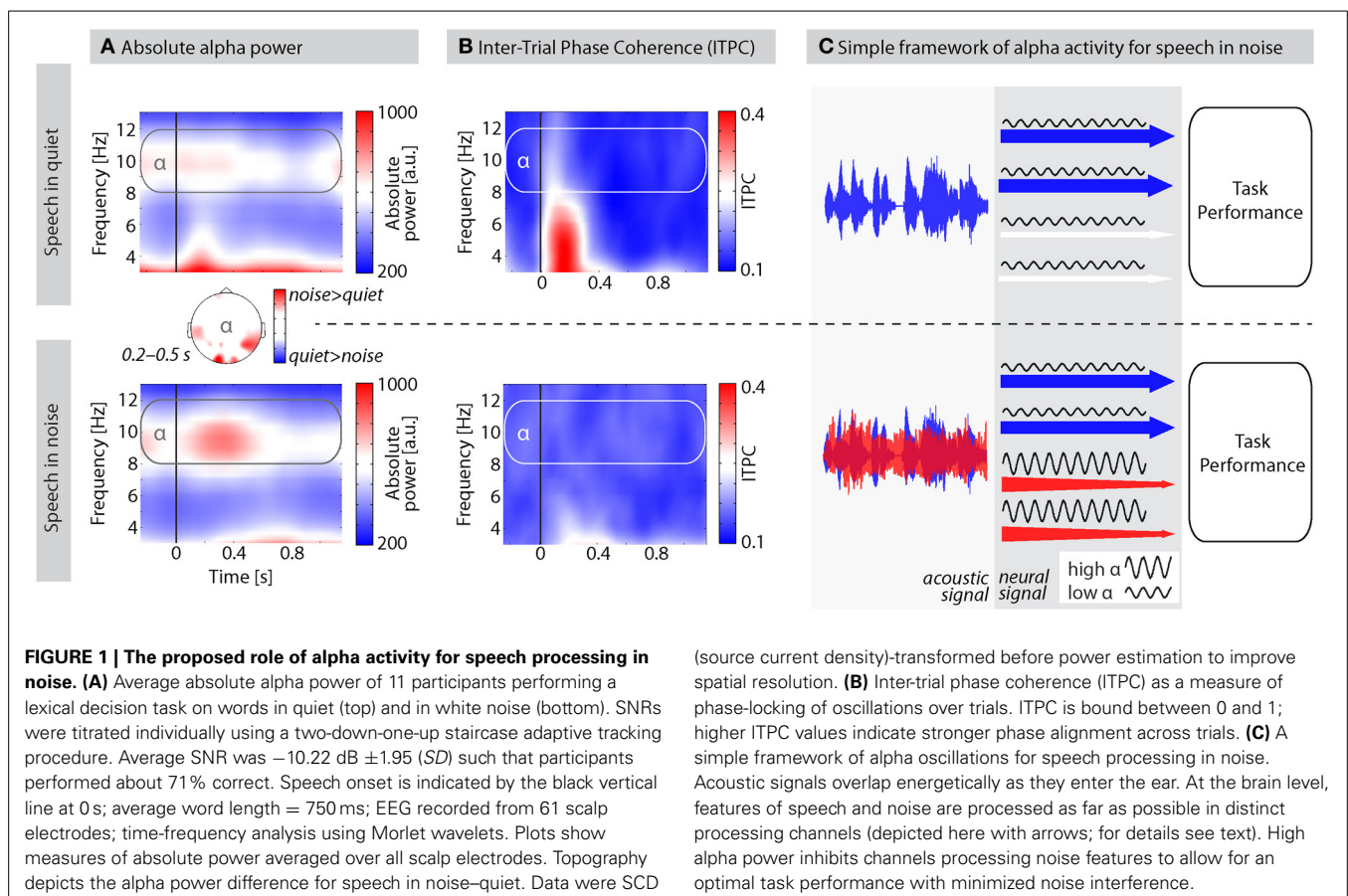


Figure 1C illustrates a tentative framework for how alpha oscillations could support auditory selective inhibition. Sounds arriving at the listener's ear must be further processed in the brain to extract task-relevant information. One way to think about the proposed mechanism is in terms of auditory object selection which requires object formation in the first place (Shinn-Cunningham, 2008). An auditory object might be formed on the basis of common spectro-temporal features, harmonicity, simultaneous onsets, or spatial grouping (Griffiths and Warren, 2004; Bizley and Cohen, 2013). We refer to all these different features used to form auditory objects as “channels” of auditory information represented by the arrows in **Figure 1C**. The concept of channels has a long tradition (Broadbent, 1958) and is inspired by the most clear distinction of target and distractor used in many dichotic listening paradigms where left and right ear channel need to be separated. Nevertheless, channels in our framework should be conceived as functional auditory processing units rather than anatomical pathways. As soon as these channels are defined, attention or inhibition can be selectively applied, given attentionally flexible fields in the auditory cortices (Petkov et al., 2004). Note that even though in the visual modality claims about alpha oscillations in feature-based (Romei et al., 2012) and object-based (Kinsey et al., 2011) attention have been made, we do not make any assumption about this distinction in our framework and use the term “channels” for both features and objects, or early and late selection.

If speech is presented in quiet (**Figure 1C**, top panel), alpha power is low in channels processing features of the speech signal to support processing of task-relevant information. Accordingly, the net resulting alpha power in the M/EEG would continue on baseline level (**Figure 1A**) and decrease during word integration (>400 ms). If, however, speech is presented in the presence of maskers (e.g., environmental noise, distracting talkers; **Figure 1C**, bottom panel), alpha power needs to be up-regulated first in those channels processing noise features before it is going to be suppressed during word integration (**Figure 1A**). Enhanced alpha activity inhibits processing of noise and thereby “protects” (Klimesch, 1999; Roux and Uhlhaas, 2014) the task- or performance-relevant information in the speech signal from noise interference.

Importantly, the up-regulation of alpha power in channels that process noise is not an automatic (“bottom-up”) process but critically depends on “top-down” attentional control. For instance, in a multi-talker situation, target and distracting talker switch roles permanently, as the listener decides to change the conversational partner. In such a situation, M/EEG alpha power would be constantly at a high level; however, the deployment of alpha power onto the different processing channels would be changing continuously.

What is the functional role of high alpha activity for word processing in noise? To answer this question, it is essential to distinguish between interpretations in which alpha activity is related to target processing from these related to noise processing. It is possible that the reduced intelligibility of words in noise leads to sub-optimal word processing and thus to less alpha suppression in brain areas relevant for speech processing (Strauß et al., 2014). The inverse mechanism, as we put forward in the current

framework, is equally likely by which alpha power is enhanced for temporarily irrelevant information and thereby compensates for perceived cognitive effort (increased when listening to speech in noise: Larsby et al., 2005; Helfer et al., 2010; Zekveld et al., 2011). In this regard, alpha would “protect” the lexical processes from noise interference. The challenge will be to experimentally dissect these (not mutually exclusive) mechanisms. We now review initial evidence for alpha's inhibitory role in audition.

Currently, there are only few studies that show alpha power modulations when participants simultaneously listen to two auditory streams, that is, one signal and one masker. In one study by Kerlin et al. (2010), participants were simultaneously listening to two spatially separated speech streams. On each trial, an initial visual cue indicated whether they were supposed to attend the left or right stream. During speech presentation, EEG alpha power was enhanced over the cerebral hemisphere contralateral to the masker, while alpha power was reduced contralateral to the to-be-attended stream. The authors concluded that this alpha lateralization indexes the direction of auditory attention to speech in space. Importantly, this finding corroborates our view that enhanced alpha power in brain areas engaged in distractor processing decreases further processing of the distractor and hence, facilitates processing of the target signal. However, two questions arise from this study: First, as the direction of auditory attention was cued visually in this study, it might be that the alpha lateralization indicates the allocation of supramodal rather than auditory selective attention (Farah et al., 1989). Second, spatial attention may play a special role not least because of auditory processing models suggesting separate what- and where-pathways (Rauschecker and Scott, 2009).

In three other recent studies, alpha power modulations were consistently found during the anticipation of auditory target signals from the left or right (Banerjee et al., 2011; Müller and Weisz, 2012; Ahveninen et al., 2013). In these studies, participants were cued to attend either the auditory event on the left or right, and to ignore the distractor on the other side. Alpha power was enhanced during the anticipation of auditory stimulation contralateral to the distractor. These results demonstrate alpha lateralization effects already during the preparation for an auditory selective listening task. This is in line with studies reporting high pre-stimulus alpha power when participants are about to miss a (visual) target (van Dijk et al., 2008; Busch et al., 2009; Romei et al., 2010). In terms of our framework (**Figure 1C**), anticipatory high alpha power successfully blocks in-depth processing of sensory information that might lead to missing the target.

However, interpretations of these studies are limited for our model, since alpha power modulations were found only during the anticipation but not during the actual processing of competing auditory streams. More data are clearly needed on the peri-stimulus alpha dynamics. As the spatial resolution of M/EEG is limited, prospective experiments could induce alpha oscillations over specific brain areas using transcranial alternating current stimulation (tACS) to assess the influence of alpha modulations on listening success under adverse acoustic conditions. Moreover, future studies could record the electrocorticogram (ECoG) directly from the cortical surface to track alpha sources and reveal the interplay between frequency bands. Such

higher spatial resolution would allow to differentiate between alpha activity in brain regions associated with processing the masker or the signal. As of now, we are left to speculate how spatially specific alpha oscillations might operate, for example along a cochleotopic gradient in primary auditory cortex. The best data to infer from stems from visual cortex, where for example Buffalo and colleagues recorded with two electrode tips in attended vs. non-attended receptive fields less than a millimeter apart and report attention-dependent opposing, and deep-layer-specific alpha changes (expressed as alpha spike-field coherence; Buffalo et al., 2011). Comparable data are, to our knowledge, still missing for auditory areas.

In the next two sections, we will elaborate first, at which levels of auditory processing alpha power might be deployed for the inhibition of different kinds of auditory maskers, and second, how age and hearing loss might affect auditory selective inhibition.

4. MASKING RELEASE VIA ALPHA ENHANCEMENT ALONG THE AUDITORY PATHWAY

So far, we have shown that alpha oscillations are an attractive neural candidate mechanism of selective auditory inhibition. There are different aspects which need to be systematically investigated in order to determine the role of alpha: Which neural circuits “deploy” or trigger high-alpha states? And in terms of the current framework: What kind of channels can be attenuated by enhanced alpha power?

Currently, there are few studies mapping the sources of alpha power during masked auditory processing. Some evidence has accumulated showing noise-invariant representations of the signal in auditory cortices (Chang et al., 2010; Ding and Simon, 2012) with the degree of invariance increasing from peripheral to cortical processing stages (Rabinowitz et al., 2013). If we assume that alpha is an important central mechanism to inhibit various types of maskers, these studies suggest that masking release via alpha enhancement might occur as early as in primary auditory cortex. A first direct hint to this idea might be the case of an illusory sound percept like tinnitus, which can be centrally suppressed by means of increasing alpha power in primary auditory cortex (Leske et al., 2013; Weisz et al., 2014). This is in line with research showing that attention modulates activity in sensory cortices corresponding to the modality of the stimulus (e.g., Heinrich et al., 2011; Wild et al., 2012). Thus, alpha activity in primary auditory cortex might be crucially contributing to inhibiting the formation of auditory objects.

In future studies investigating underlying alpha sources, a distinction between energetic and informational masking might be crucial (Brungart et al., 2001; Mattys et al., 2009; Scott and McGettigan, 2013; for a more comprehensive overview of potential adverse listening conditions see Mattys et al., 2012). Energetic masking describes the competition of auditory target and masker in the auditory periphery due to spectro-temporal overlap of the two signals, causing an overlap of excitation patterns in the cochlea and auditory nerve (Durlach et al., 2003). One type of background signal often assumed to cause primarily energetic masking is white noise (e.g., Arbogast et al., 2005) which is quasi-stationary and has high energy in a broad frequency range (for discussion see Stone et al., 2012). Although informational

masking is sometimes defined only negatively as all masking effects not accounted for by energetic masking (cf. Gutschalk et al., 2008), a more refined definition is required, especially when it comes to speech processing. When target speech is masked by a competing talker, it is not just the energetic overlap of the two signals that causes masker interference. Rather, the speech masker initiates phonetic and semantic processing that interferes with the linguistic processing of the target (Schneider et al., 2007). Thus, informational masking describes the interference of target and masker at a more central, cognitive level, whereas energetic masking refers to energetic overlap in the auditory periphery.

According to the framework described above, alpha oscillations might be important for inhibition of both types of maskers, however, in different brain areas. We presume that energetic maskers are inhibited by enhanced alpha activity in auditory cortex (Müller and Weisz, 2012). In contrast, processing of informational maskers like competing speech should rather be inhibited by alpha activity in higher auditory areas such as posterior superior temporal gyrus (pSTG) and beyond, relevant for linguistic processing (Scott et al., 2004, 2009). In addition to the proposed inhibition of auditory input, alpha oscillations are involved in supramodal or crossmodal inhibition of the currently task-irrelevant modality (Banerjee et al., 2011).

5. EFFECTS OF AGE AND HEARING LOSS ON AUDITORY DISTRACTOR INHIBITION

In acoustically demanding multi-talker situations, older listeners typically experience more difficulties compared with younger adults. It is however unclear, in how far these difficulties are caused by age-related decline in perceptual auditory acuity (hearing loss or loss of temporal and spectral resolution; Fostick and Babkoff, 2013), decline of cognitive functioning with age, or both (Wingfield et al., 2005). Crucial for the present framework, however, both auditory perceptual and cognitive decline could lead to insufficient masker inhibition. First, compared with normal-hearing controls, listeners with hearing loss are less successful in utilizing spectral (Lorenzi et al., 2006), temporal (Tremblay et al., 2003), and spatial auditory cues (Neher et al., 2009) important for the perceptual segregation of different sound sources. Thus, attending to relevant and inhibiting irrelevant sound sources is impaired, as auditory features are lacking to distinguish the different sound sources in the first place (Shinn-Cunningham and Best, 2008). Second, age negatively affects many aspects of cognitive functioning (Park et al., 2003), amongst it the ability to suppress irrelevant but salient auditory distractors (Chao and Knight, 1997; Tun et al., 2002; Passow et al., 2014). Thus, even if the perceptual segregation of sound sources is accomplished successfully, the insufficient inhibition of maskers may cause interference.

In line with prior studies that found age effects on brain oscillatory activity in the alpha frequency range (Yordanova et al., 1998; Klimesch, 1999; Böttger et al., 2002), we consider it valuable to investigate alpha oscillations in demanding listening tasks as an indicator of age-dependent auditory cognitive effort of masker inhibition. We presume that auditory selective inhibition, realized by alpha activity in channels relevant for masker processing (Figure 1C), serves as a compensatory mechanism as multi-talker listening conditions become more demanding, for instance due

to a decreasing signal-to-noise ratio (SNR). The study of alpha oscillations could help to reveal how listeners of different age exert top-down attentional control to facilitate processing of task-relevant signals and inhibit processing of interfering maskers. In particular, this line of research might foster the understanding of why older listeners find it more exhausting to participate in cocktail party-like listening situations compared with younger listeners (Pichora-Fuller, 2003).

6. CONCLUSIONS

In this perspective article, we have presented a framework for studying alpha oscillations as a tool for auditory selective inhibition in challenging listening situations. We have presented initial evidence qualifying alpha oscillations as a pivotal mechanism affecting listening in multi-talker situations. Future studies could expand these findings and study the role of alpha oscillations (1) during speech perception in ecologically valid listening situations, (2) in the presence of energetic and informational maskers, and (3) for aging and hearing-impaired listeners.

ACKNOWLEDGMENTS

Antje Strauß, Malte Wöstmann, and Jonas Obleser are supported by a Max Planck Research Grant to Jonas Obleser. The authors are grateful for in-depth discussions with the members of the Max Planck Research Group “Auditory Cognition” during manuscript preparation.

REFERENCES

- Ahveninen, J., Huang, S., Belliveau, J. W., Chang, W.-T., and Hämäläinen, M. (2013). Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. *J. Cogn. Neurosci.* 25, 1926–1943. doi: 10.1162/jocn_a_00452
- Arbogast, T. L., Mason, C. R., and Kidd, G. Jr. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 117, 2169–2180. doi: 10.1121/1.1861598
- Banerjee, S., Snyder, A. C., Molholm, S., and Foxe, J. J. (2011). Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supramodal or sensory-specific control mechanisms? *J. Neurosci.* 31, 9923–9932. doi: 10.1523/JNEUROSCI.4660-10.2011
- Berger, H. (1931). Über das Elektroencephalogramm des Menschen. *Arch. Psychiatr. Nervenkr.* 94, 16–60. doi: 10.1007/BF01835097
- Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565
- Böttger, D., Herrmann, C. S., and von Cramon, D. Y. (2002). Amplitude differences of evoked alpha and gamma oscillations in two different age groups. *Int. J. Psychophysiol.* 45, 245–251. doi: 10.1016/S0167-8760(02)00031-4
- Broadbent, D. E. (1958). *Perception and Communication*. Oxford, UK: Pergamon Press. doi: 10.1037/10037-000
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J. Acoust. Soc. Am.* 110, 2527–2538. doi: 10.1121/1.1408946
- Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11262–11267. doi: 10.1073/pnas.1011284108
- Busch, N. A., Dubois, J., and VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876. doi: 10.1523/JNEUROSCI.0113-09.2009
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432. doi: 10.1038/nn.2641
- Chao, L. L., and Knight, R. T. (1997). Prefrontal deficits in attention and inhibitory control with aging. *Cereb. Cortex* 7, 63–69. doi: 10.1093/cercor/7.1.63
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Cohen, M. X., Elger, C. E., and Fell, J. (2009). Oscillatory activity and phase-amplitude coupling in the human medial frontal cortex during decision making. *J. Cogn. Neurosci.* 21, 390–402. doi: 10.1162/jocn.2008.21020
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222. doi: 10.1146/annurev.ne.18.030195.001205
- Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109
- Durlach, N. I., Mason, C. R. Jr., G. K., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). Note on informational masking (I). *J. Acoust. Soc. Am.* 113, 2984–2987. doi: 10.1121/1.1570435
- Farah, M. J., Wong, A. B., Monheit, M. A., and Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: modality-specific or supramodal? *Neuropsychologia* 27, 461–470. doi: 10.1016/0028-3932(89)90051-1
- Fostick, L., and Babkoff, H. (2013). Temporal and non-temporal processes in the elderly. *J. Basic Clin. Physiol. Pharmacol.* 24, 191–199. doi: 10.1515/jbcpp-2013-0049
- Foxe, J. J., and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* 2:154. doi: 10.3389/fpsyg.2011.00154
- Freunberger, R., Fellinger, R., Sauseng, P., Gruber, W., and Klimesch, W. (2009). Dissociation between phase-locked and nonphase-locked alpha oscillations in a working memory task. *Hum. Brain Mapp.* 30, 3417–3425. doi: 10.1002/hbm.20766
- Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention-focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 437–455. doi: 10.1016/j.conb.2007.07.011
- Griffiths, T. D., and Warren, J. D. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892. doi: 10.1038/nrn1538
- Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* 6:e138. doi: 10.1371/journal.pbio.0060138
- Haegens, S., Luther, L., and Jensen, O. (2012). Somatosensory anticipatory alpha activity increases to suppress distracting input. *J. Cogn. Neurosci.* 24, 677–685. doi: 10.1162/jocn_a_00164
- Haegens, S., Nacher, V., Luna, R., Romo, R., and Jensen, O. (2011). α -oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19377–19382. doi: 10.1073/pnas.1117190108
- Hartmann, T., Schlee, W., and Weisz, N. (2012). It's only in your head: expectancy of aversive auditory stimulation modulates stimulus-induced auditory cortical alpha desynchronization. *Neuroimage* 60, 170–178. doi: 10.1016/j.neuroimage.2011.12.034
- Heinrich, A., Carlyon, R. P., Davis, M. H., and Johnsrude, I. S. (2011). The continuity illusion does not depend on attentional state: fMRI evidence from illusory vowels. *J. Cogn. Neurosci.* 23, 2675–2689. doi: 10.1162/jocn.2011.21627
- Helfer, K. S., Chevalier, J., and Freyman, R. L. (2010). Aging, spatial cues, and single- versus dual-task performance in competing speech perception. *J. Acoust. Soc. Am.* 128, 3625–3633. doi: 10.1121/1.3502462
- Jensen, O., Bonnefond, M., and VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn. Sci.* 16, 200–206. doi: 10.1016/j.tics.2012.03.002
- Jensen, O., Gelfand, J., Kounios, J., and Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cereb. Cortex* 12, 877–882. doi: 10.1093/cercor/12.8.877
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Kaiser, J., Heidegger, T., Wibral, M., Altmann, C. F., and Lutzenberger, W. (2007). Alpha synchronization during auditory spatial short-term memory. *Neuroreport* 18, 1129–1132. doi: 10.1097/WNR.0b013e32821c553b
- Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *J. Neurosci.* 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010
- Kinsey, K., Anderson, S. J., Hadjipapas, A., and Holliday, I. E. (2011). The role of oscillatory brain activity in object processing and figure-ground

- segmentation in human vision. *Int. J. Psychophysiol.* 79, 392–400. doi: 10.1016/j.ijpsycho.2010.12.007
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007
- Larsby, B., Hällgren, M., Lyxell, B., and Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *Int. J. Audiol.* 44, 131–143. doi: 10.1080/14992020500057244
- Lee, A. K. C., Larson, E., Maddox, R. K., and Shinn-Cunningham, B. G. (2013). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hear. Res.* 307, 111–120. doi: 10.1016/j.heares.2013.06.010
- Leiberg, S., Lutzenberger, W., and Kaiser, J. (2006). Effects of memory load on cortical oscillatory activity during auditory pattern working memory. *Brain Res.* 1120, 131–140. doi: 10.1016/j.brainres.2006.08.066
- Leske, S., Tse, A., Oosterhof, N. N., Hartmann, T., Müller, N., Keil, J., et al. (2013). The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. *Neuroimage* 88C, 69–78. doi: 10.1016/j.neuroimage.2013.11.014
- Lorenz, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18866–18869. doi: 10.1073/pnas.0607364103
- Luo, H., Husain, F. T., Horwitz, B., and Poeppel, D. (2005). Discrimination and categorization of speech and non-speech sounds in an MEG delayed-match-to-sample study. *Neuroimage* 28, 59–71. doi: 10.1016/j.neuroimage.2005.05.040
- Mathewson, K. E., Lleras, A., Beck, D. M., Fabiani, M., Ro, T., and Gratton, G. (2011). Pulsed out of awareness: EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. *Front. Psychol.* 2:99. doi: 10.3389/fpsyg.2011.00099
- Mattys, S. L., Brooks, J., and Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cogn. Psychol.* 59, 203–243. doi: 10.1016/j.cogpsych.2009.04.001
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Lang. Cogn. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006
- Melara, R. D., Rao, A., and Tong, Y. (2002). The duality of selection: excitatory and inhibitory processes in auditory selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 279–306. doi: 10.1037/0096-1523.28.2.279
- Meyer, L., Obleser, J., and Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex* 49, 711–721. doi: 10.1016/j.cortex.2012.03.006
- Müller, N., and Weisz, N. (2012). Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. *Cereb. Cortex* 22, 1604–1613. doi: 10.1093/cercor/bhr232
- Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., et al. (2009). Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: effects of hearing loss, age, and cognition. *Int. J. Audiol.* 48, 758–774. doi: 10.3109/14992020903079332
- Obleser, J., and Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cereb. Cortex* 22, 2466–2477. doi: 10.1093/cercor/bhr325
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* 32, 12376–12383. doi: 10.1523/JNEUROSCI.4908-11.2012
- Palva, S., and Palva, J. M. (2011). Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Front. Psychol.* 2:204. doi: 10.3389/fpsyg.2011.00204
- Park, H. L., O'Connell, J. E., and Thomson, R. G. (2003). A systematic review of cognitive decline in the general elderly population. *Int. J. Geriatr. Psychiatry* 18, 1121–1134. doi: 10.1002/gps.1023
- Passow, S., Westerhausen, R., Hugdahl, K., Wartenburger, I., Heekeren, H. R., Lindenberger, U., et al. (2014). Electrophysiological correlates of adult age differences in attentional control of auditory processing. *Cereb. Cortex* 24, 249–260. doi: 10.1093/cercor/bhs306
- Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., and Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nat. Neurosci.* 7, 658–663. doi: 10.1038/nn1256
- Pfurtscheller, G., Stancák Jr., A., and Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band – an electrophysiological correlate of cortical idling: a review. *Int. J. Psychophysiol.* 24, 39–46. doi: 10.1016/S0167-8760(96)00066-9
- Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *Int. J. Audiol.* 42, 2S26–2S32. doi: 10.3109/14992020309074641
- Rabinowitz, N. C., Willmore, B. D. B., King, A. J., and Schnupp, J. W. H. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol.* 11:e1001710. doi: 10.1371/journal.pbio.1001710
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- Romei, V., Gross, J., and Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? *J. Neurosci.* 30, 8692–8697. doi: 10.1523/JNEUROSCI.0160-10.2010
- Romei, V., Thut, G., Mok, R. M., Schyns, P. G., and Driver, J. (2012). Causal implication by rhythmic transcranial magnetic stimulation of alpha frequency in feature-based local vs. global attention. *Eur. J. Neurosci.* 35, 968–974. doi: 10.1111/j.1460-9568.2012.08020.x
- Roux, F., and Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010
- Roux, F., Wibral, M., Singer, W., Aru, J., and Uhlhaas, P. J. (2013). The phase of thalamic alpha activity modulates cortical gamma-band activity: evidence from resting-state MEG recordings. *J. Neurosci.* 33, 17827–17835. doi: 10.1523/JNEUROSCI.5778-12.2013
- Schneider, B. A., Li, L., and Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *J. Am. Acad. Audiol.* 18, 559–572. doi: 10.3766/jaaa.18.7.4
- Scott, S. K., and McGettigan, C. (2013). The neural processing of masked speech. *Hear. Res.* 303, 58–66. doi: 10.1016/j.heares.2013.05.001
- Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., and Wise, R. J. S. (2009). The neural processing of masked speech: evidence for different mechanisms in the left and right temporal lobes. *J. Acoust. Soc. Am.* 125, 1737–1743. doi: 10.1121/1.3050255
- Scott, S. K., Rosen, S., Wickham, L., and Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J. Acoust. Soc. Am.* 115, 813–821. doi: 10.1121/1.1639336
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci.* 12, 182–186. doi: 10.1016/j.tics.2008.02.003
- Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306
- Spaak, E., Bonnefond, M., Maier, A., Leopold, D. A., and Jensen, O. (2012). Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. *Curr. Biol.* 22, 2313–2318. doi: 10.1016/j.cub.2012.10.020
- Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *J. Acoust. Soc. Am.* 132, 317–326. doi: 10.1121/1.4725766
- Strauß, A., Kotz, S. A., Scharinger, M., and Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage*. doi: 10.1016/j.neuroimage.2014.04.005. [Epub ahead of print].
- Strauss, D. J., Corona-Strauss, F. I., Trenado, C., Bernarding, C., Reith, W., Latzel, M., et al. (2010). Electrophysiological correlates of listening effort: neurodynamical modeling and measurement. *Cogn. Neurodyn.* 4, 119–131. doi: 10.1007/s11571-010-9111-3
- Tremblay, K. L., Piskosz, M., and Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clin. Neurophysiol.* 114, 1332–1343. doi: 10.1016/S1388-2457(03)00114-7
- Tun, P. A., O'Kane, G., and Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychol. Aging* 17, 453–467. doi: 10.1037/0882-7974.17.3.453
- van Dijk, H., Schoffelen, J.-M., Oostenveld, R., and Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *J. Neurosci.* 28, 1816–1823. doi: 10.1523/JNEUROSCI.1853-07.2008

- Weisz, N., Lüchinger, C., Thut, G., and Müller, N. (2014). Effects of individual alpha rTMS applied to the auditory cortex and its implications for the treatment of chronic tinnitus. *Hum. Brain Mapp.* 35, 14–29. doi: 10.1002/hbm.22152
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012
- Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., and Obleser, J. (2014). Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cereb. Cortex*. doi: 10.1093/cercor/bhu004. [Epub ahead of print].
- Wingfield, A., Tun, P. A., and McCoy, S. L. (2005). Hearing loss in older adulthood what it is and how it interacts with cognitive performance. *Curr. Direct. Psychol. Sci.* 14, 144–148. doi: 10.1111/j.0963-7214.2005.00356.x
- Yordanova, J. Y., Kolev, V. N., and Başar, E. (1998). EEG theta and frontal alpha oscillations during auditory processing change with aging. *Electroencephalogr. Clin. Neurophysiol.* 108, 497–505. doi: 10.1016/S0168-5597(98)00028-8
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear Hear.* 32, 498–510. doi: 10.1097/AUD.0b013e31820512bb
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 February 2014; accepted: 08 May 2014; published online: 28 May 2014.

Citation: Strauß A, Wöstmann M and Obleser J (2014) Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8:350. doi: 10.3389/fnhum.2014.00350

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Strauß, Wöstmann and Obleser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Induction of plasticity in the human motor cortex by pairing an auditory stimulus with TMS

Paul F. Sowman^{1,2,3*}, Søren S. Dueholm^{1,4}, Jesper H. Rasmussen^{1,4} and Natalie Mrachacz-Kersting⁴

¹ Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia

² Perception and Action Research Centre (PARC), Faculty of Human Sciences, Macquarie University, Sydney, NSW, Australia

³ Australian Research Council Centre of Excellence in Cognition and its Disorders (CCD), Macquarie University, Sydney, NSW, Australia

⁴ Department of Health Science and Technology, Center for Sensory-Motor Interaction (SMI), Aalborg University, Aalborg, Denmark

Edited by:

Carolyn McGettigan, Royal Holloway University of London, UK

Reviewed by:

Alessandro D'Ausilio, Italian Institute of Technology, Italy
Dan Kennedy-Higgins, University College London–Speech, Hearing and Phonetic Sciences, UK

*Correspondence:

Paul F. Sowman, Department of Cognitive Science, Level 3 Australian Hearing Hub, 16 University Drive, Macquarie University, Sydney, NSW 2109, Australia
e-mail: paul.sowman@mq.edu.au

Acoustic stimuli can cause a transient increase in the excitability of the motor cortex. The current study leverages this phenomenon to develop a method for testing the integrity of auditorimotor integration and the capacity for auditorimotor plasticity. We demonstrate that appropriately timed transcranial magnetic stimulation (TMS) of the hand area, paired with auditorily mediated excitation of the motor cortex, induces an enhancement of motor cortex excitability that lasts beyond the time of stimulation. This result demonstrates for the first time that paired associative stimulation (PAS)-induced plasticity within the motor cortex is applicable with auditory stimuli. We propose that the method developed here might provide a useful tool for future studies that measure auditory-motor connectivity in communication disorders.

Keywords: paired associative stimulation, transcranial magnetic stimulation, auditory motor integration, speech sounds, plasticity, motor cortex, auditory cortex

INTRODUCTION

Paired associative stimulation (PAS) is a technique used to experimentally induce long-lasting changes in cortical excitability (Stefan et al., 2002; Ridding and Flavel, 2006; Mrachacz-Kersting et al., 2007; Murakami et al., 2008; Kumpulainen et al., 2012). Most commonly in PAS studies, electrical stimulation of the median nerve is paired with transcranial magnetic stimulation (TMS) of the contralateral motor cortex. The nerve impulse resulting from the somatosensory stimulus can be timed to arrive at the cortical level milliseconds prior to the TMS pulse in order to induce a long-lasting increase in excitability—a process that is thought to be mediated by a Hebbian long-term potentiation (LTP)-like process (Stefan et al., 2002).

In recent years, modified PAS protocols have been designed that apply more ecologically valid stimuli in place of either the TMS or the electrical somatosensory stimulation, e.g., TMS paired with movement (Thabit et al., 2010) or electrical somatosensory stimulation paired with motor imagery (Mrachacz-Kersting et al., 2012). PAS protocols have also moved beyond ubiquitous sensorimotor associations to demonstrate that pairing a TMS-induced cortical activation outside the motor cortex with a homotopic sensory activation can induce enhanced responses to sensory inputs. For example, Schecklmann et al. (2011) showed that pairing a TMS pulse to the auditory cortex with a simple tone could induce a prolonged decrement of the auditory evoked potential. Cortical stimulation has also been paired with visual stimuli to demonstrate the capacity for

visuomotor integration to mediate plastic changes in motor cortex (Suppa et al., 2013). To date however, the connections known to exist between the auditory and motor domains have not been tested for their capacity to induce motor cortex plasticity.

A number of well-described functional links between audition and the motor system exist. These range from protective reflexive motor activations in response to signals of potential danger (Forbes and Sherrington, 1914) to the complex feedback and feedforward communication necessary for fluent speech to occur (Tourville et al., 2008; Perkell, 2012). These connections allow us to, for example, modulate the volume of our speech to appropriately match the ambient environmental noise (Lane and Tranel, 1971) or modulate the sensitivity of our sensory system to compensate for speech-induced refference (Curio et al., 2000).

Motoric activation via auditory inputs has been demonstrated in a number of experiments that have used TMS to probe the link between speech perception and motor representations. Modulation of motor cortical excitability during speech perception has been demonstrated to occur in the cortical representations of the hand (Flöel et al., 2003), lips (Watkins et al., 2003) and tongue (Fadiga et al., 2002; Roy et al., 2008).

Despite evidence suggesting a strong connection between auditory and motor centers, auditory stimuli have not yet been used in a modified PAS study to induce plasticity in the motor area. The aim of the current study was to investigate whether it is possible to induce plasticity in the motor system by pairing auditory stimuli and TMS. The development of such a protocol would in future allow for the direct investigation of auditorimotor

linkages in a number of disorders where these are thought to be abnormal. Auditorimotor disconnection or dysfunction has, for example, been proposed to underpin the speech dysfluencies that characterize stuttering (Neef et al., 2011) and the misattribution of self-produced speech that may produce auditory hallucinations in schizophrenia (Ford et al., 2005).

METHODS

Two separate experiments (A and B) were conducted. Given that the timing of stimuli in a PAS protocol is critical for facilitating plastic change (Stefan et al., 2002; Wolters et al., 2005; Mrachacz-Kersting et al., 2007; Murakami et al., 2008; Kumpulainen et al., 2012), the aim of Experiment A was to find, at a group level, the optimal offset timing of the motor cortical excitation from the onset of the auditory stimulus. This offset was determined by applying TMS pulses at different latencies relative to the onset of the auditory stimulus and measuring the conditioned motor evoked potential (MEP) in the right first dorsal interosseus (FDI) muscle.

The PAS protocol implemented in Experiment B was informed by the results of Experiment A. First a baseline session was conducted where MEPs (TMS with no auditory stimuli) were collected and saved as pre-PAS measurements. This was followed by an intervention block which consisted of the auditorimotor PAS-protocol. During the intervention block, subjects received an auditory stimulus paired with TMS using the optimal time latency between stimulations that was found in Experiment A. After the intervention session, post-PAS MEPs were recorded immediately after and then 15 min after the session ended (post and post15, respectively). By comparing pre- with post-MEPs and post15-MEPs it was possible to evaluate whether motor cortex excitability changes had occurred, how fast they evolved and whether they were long-lasting.

EXPERIMENT A—TIMING OF STIMULI

SUBJECTS

Experiment A was performed on 12 healthy right-handed volunteers (9 males), aged 18–36 years (mean 24.2 ± 5.0 years). Prior to commencement of the experiment subjects completed a standard TMS screening questionnaire and provided written informed consent. None of the subjects reported any history of hearing impairment, neurological disease or mental illness, was taking regular medication or had a history brain injury. This study was reviewed and approved by the Human Research Ethics Committee of Macquarie University.

EXPERIMENTAL PROCEDURE

Subjects were seated in a chair with their right arm and hand resting in a comfortable position on an armrest. An armrest was used in order to eliminate hand movements during recordings. During the experiment the subject was told to relax, avoid any movement of the right arm and hand and to have their eyes open. Surface EMG (sEMG) was recorded ($1000 \times$ gain, bandpass filtered from 20–500 Hz) from a bipolar electrode (Medi-Trace 100, Kendall/Tyco Healthcare, USA) montage. One electrode was placed over the muscle belly of the right FDI muscle and the other

electrode was placed over the proximal metacarpal of the index finger.

A monophasic transcranial magnetic stimulator (Magstim model 200, Magstim, Whitland, UK), with a focal figure-of-eight stimulating coil (90-mm outer diameter), was used to elicit MEPs from the right FDI muscle. The stimulating coil was held tangentially to the skull with the coil oriented 45° to the parasagittal plane and the handle pointing laterally and posteriorly. The center of the coil junction was placed over the primary motor cortex (M1) hand area of the left hemisphere and the “motor hot spot” was determined as the site where TMS consistently elicited the largest MEPs.

Resting motor threshold (MT) was determined by finding the lowest stimulation intensity of the motor hotspot for the right FDI needed in order to obtain an MEP with a peak-to-peak amplitude of $50 \mu\text{V}$ in 5 out of 10 consecutive stimulations. The TMS test intensity was then set at 120% of resting MT. Eight different TMS conditions were tested. These consisted of seven auditory-stimulation/TMS pairs and one TMS condition without associated auditory stimulation (baseline). The auditory-stimulation/TMS pairs consisted of a test TMS pulse applied at one of seven different intervals (25, 50, 100, 150, 200, 250 and 300 ms) after the onset of the auditory stimulus. The auditory stimulus consisted of a male voice pronouncing the word “Hey!” played back at 80 dB SPL via Etymotic ER-1 insert tube-phones. We chose to use a speech sounds because previous research suggests that speech sounds strongly activate the motor cortex e.g., Flöel et al. (2003). However, other evidence suggests that the motor cortex might be also activated by non speech sounds (Watkins et al., 2003; Alibiglou and Mackinnon, 2012) so we also included a condition in which the auditory stimulus matched the amplitude envelope of the speech stimulus but consisted entirely of white noise (Pulvermüller et al., 2006). This signal-correlated noise (SCN) stimulus was created using Praat (Boersma and Weenink, 2013). Time and frequency domain comparisons of the two signals are displayed in Figure 1.

The order of all seven auditory-stimulation/TMS pairs and stimulus types (speech or SCN) was randomly intermingled and presented with an intertrial interval (ITI) that randomly varied between 4000 and 5000 ms in two blocks such that the total number of stimuli per condition was 16. The total number of trials was hence 128 (16 baseline trials + 7×16 conditioned trials). The duration of the experiment was approximately 25 min.

DATA PROCESSING

Offline MEP analysis was conducted using a custom MATLAB (The Mathworks, USA) script. The average MEP amplitude calculated for each sound type and auditory-stimulation/TMS pair was expressed as a function of the average pre MEP (baseline).

STATISTICAL ANALYSIS

A repeated measures ANOVA with the factors delay (auditory-stimulus/TMS interval) and condition (speech or SCN) was performed on the averaged MEPs. A two-tailed, one-sample *t*-test was then used to determine the time points at which the conditioned MEPs differed significantly from baseline using an α -value of 0.05.

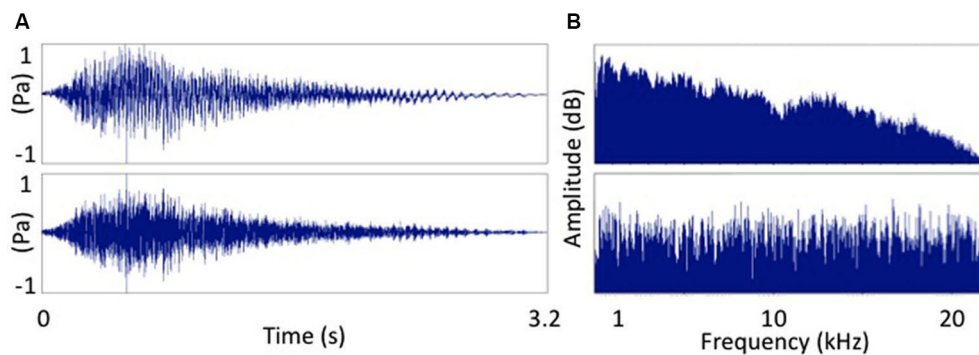


FIGURE 1 | Two sounds used as auditory stimuli. (A) The word “Hey!” and **(B)** signal correlated noise version of **(A)**. Frequency spectra of the two

auditory stimuli. **(A)** The word “Hey!” and **(B)** signal correlated noise (white noise) version of **(A)**.

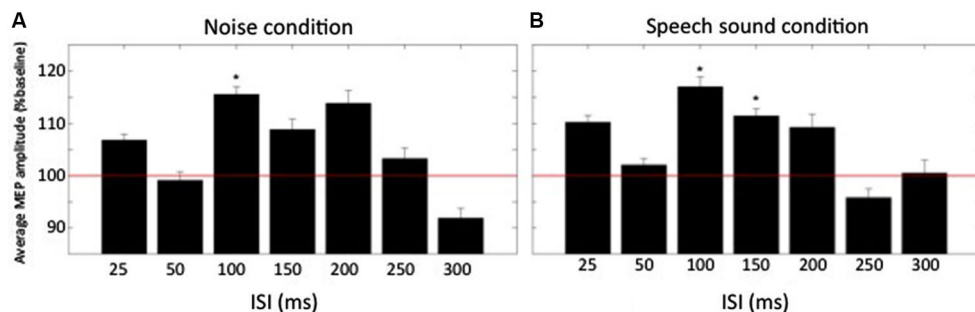


FIGURE 2 | Normalized averaged MEP amplitudes (+SEM) at different times relative to the conditioning auditory stimulus. (A) MEP amplitudes for the condition “noise” for all ISIs ($n = 12$). **(B)** MEP amplitudes for the

condition “speech sound.” Baseline is represented by the red horizontal line. * denotes average amplitude significantly different from baseline ($p < 0.05$).

EXPERIMENT B—AUDITORIMOTOR PAS

SUBJECTS

Experiment B was performed on 10 healthy right-handed volunteers (8 males), aged 18–31 years (mean 24.5 ± 3.3 years) without any prior neurological medical history. Written informed consent was obtained from each subject before participation in the study.

EXPERIMENTAL PROCEDURE

The procedure used in Experiment B was similar to the one used in Experiment A. The main difference was that a single auditory stimulus/TMS interval (100 ms) was used during the PAS induction period in Experiment B. As no difference in MEP facilitation between the speech and SCN stimulus conditions was found in Experiment A we arbitrarily chose to use only the speech stimulus in Experiment B. PAS induction following baseline MEP recording consisted of a total of 200 auditory stimulus/TMS pairs applied with a 4000–5000 ms random interval between each pair. A 2 min pause in stimulation after 100 pairs were applied was included. The total duration of the experiment was approximately 27 min (introduction: 10 min, part one: 7.5 min, pause: 2 min, part two: 7.5 min).

STATISTICAL ANALYSIS

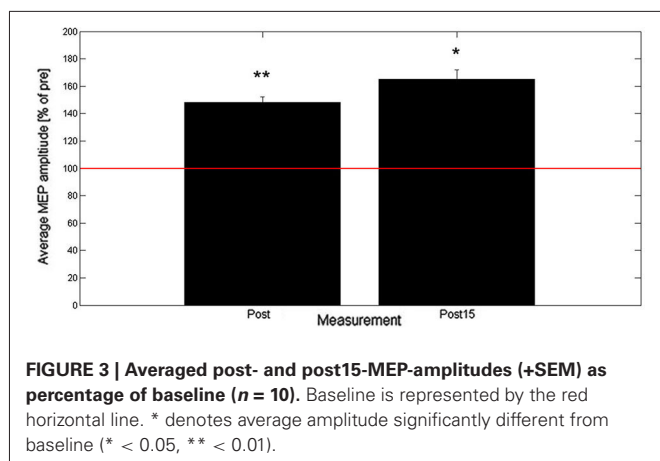
A two-tailed, one-sample t -test was used to determine significant differences between pre-MEPs (baseline), post-MEPs and post15-MEPs using an α -value of 0.05.

RESULTS

Mean (\pm SEM) MEP threshold in Experiment A was $45.5 \pm 2.1\%$ of stimulator output and $46.6 \pm 2.4\%$ in Experiment B.

Results from Experiment A are shown in **Figure 2**. A repeated measures ANOVA showed that there was a significant effect of delay on the size of the MEP $F_{(6,66)} = 2.3$, $p = 0.045$. There was no significant effect of condition nor significant interaction between delay and condition. Within condition comparison of mean normalized MEPs to baseline by means of a two-tailed one-sample t -test revealed that in the noise condition, MEPs were significantly increased above baseline for one ISI: 100 ms ($115.5 \pm 5.2\%$ of baseline, $t_{(11)} = 3.0$, $p = 0.012$). For the speech sound condition two ISIs had MEPs that were significantly increased above baseline: ISI = 100 ms ($117.0 \pm 6.5\%$ of baseline, $t_{(11)} = 2.6$, $p = 0.023$) and ISI = 150 ms ($111.4 \pm 4.7\%$ of baseline, $t_{(11)} = 2.4$, $p = 0.035$).

Results from Experiment B show that across all subjects the averaged MEP peak-to-peak amplitude increased to 148% (post)



and 165% (post15) of baseline as shown in **Figure 3**. Two-tailed one-sample t -tests showed a significant increase in normalized MEP peak-to-peak amplitude for post ($t_{(9)} = 3.8$, $p = 0.004$) and post15 ($t_{(9)} = 2.9$, $p = 0.018$). Comparison between post and post15 by means of a paired t -test revealed no significant difference ($t_{(9)} = 1.06$, $p = 0.32$).

DISCUSSION

The current study demonstrates for the first time that long-lasting motor cortical plasticity can be induced by an auditorimotor PAS paradigm. This result is significant because it not only provides a new method for investigating auditorimotor integration, but importantly, also a method to directly probe the brain's capacity for auditorimotor plasticity.

We utilized a two-stage approach in developing this PAS paradigm. First, we identified the optimal ISI for eliciting an enhanced MEP response compared to baseline. This paradigm follows the empirical approach developed by Mrachacz-Kersting et al. (2007) to investigate PAS induced plasticity in the cortical representation of tibialis anterior. The optimal interval we found fits well with the temporal structure of the auditory N1 to speech sounds which peaks 100 ms after stimulus onset e.g., Liotti et al. (2010), and agrees with the TMS findings of Fadiga et al. (2002) and those of Roy et al. (2008), who found “phonological motor resonance” was present at 100 ms after their target speech sound stimulus onsets. In both studies the authors applied TMS to the tongue motor representation following the presentation of pseudo-words containing double consonants. The MEP response that they recorded in the tongue peaked in amplitude when the auditory stimulus to TMS interval was 100 ms.

While we used a speech stimulus in these experiments, the lack of difference between the response to the speech stimulus and SCN found in Experiment A suggests that under the experimental conditions we have imposed, i.e., a repetitive presentation of a speech sound without the requirement for engagement on the part of the subject, the stimulus may not be processed as speech *per se* and should rather be considered a non-specific acoustic stimulus. This fact may explain why our results differ in part to those of Watkins et al. (2003) and Murakami et al. (2011) whose findings suggest that auditory-induced motor modulations

related to speech listening are confined to the cortical representations of those muscles involved in articulation. Indeed, there is now a significant body of evidence to support the somatotopic arrangement of speech gesture perception (Fadiga et al., 2002; Roy et al., 2008; D'Ausilio et al., 2009, 2011; Möttönen and Watkins, 2009; Sato et al., 2010) but such findings do not necessarily rule out the non-specific motor activations in response to both speech and non-speech acoustic stimuli that have been documented using both TMS and other methods (Flöel et al., 2003; Alibiglou and Mackinnon, 2012; Fujioka et al., 2012).

The current study shows that repeated pairing of an acoustic stimulus with a TMS pulse to the motor cortex representation of the hand leads to a rapidly-evolving, long-lasting increase in cortical excitability. This effect was induced with an ISI of 100 ms, a time interval that corresponded to the point of peak enhancement in the acoustic stimulus-conditioned MEP. Given that this ISI was converged upon using a method that used discrete intervals with a minimum step of 50 ms, it is expected that this PAS technique could be refined further by re-examining the optimal sound-to-TMS interval using smaller time steps (i.e., less than 25 ms) centered around 100 ms. Moreover, using auditory evoked potentials to discover individualized N1 latencies, and then using these as the basis for the PAS ISI would likely refine the technique further. Since we were able to find a significant PAS effect in this proof of concept study, we posit that auditorimotor PAS is a robust effect that will provide a powerful tool for studying auditorimotor plasticity in the future.

Auditorimotor plasticity i.e., the capacity for strengthening of auditorimotor connections within the brain is essential for the acquisition of speech and the learning of musical competence. For this reason, techniques that can probe the brain's capacity for auditorimotor plasticity provide the opportunity to investigate some of the hypothesized mechanisms of conditions such as stuttering and specific language impairment (SLI) in which disordered motor learning has been documented (Namasivayam and van Lieshout, 2008; Mayor-Dubois et al., 2014). Both of those conditions have been associated with disordered sensorimotor integration (Hill, 2001; Neef et al., 2011; Cai et al., 2012, 2014) and, in the case of SLI, with disordered auditorimotor plasticity (Kurt et al., 2012). Additionally, disorders such as schizophrenia and tinnitus have been associated with disrupted auditorimotor connections (Cacace, 2003; Ford et al., 2005; Langguth et al., 2005) and synaptic plasticity (Møller, 2003; Stephan et al., 2009); the technique described herein is therefore a novel means to assess these associations. Beyond mechanistic investigation of disorders, associative stimulation using TMS has also been proposed as a therapeutic modality (Uy et al., 2003; Jayaram and Stinear, 2008; Michou et al., 2013). If it is established that disorders such as those described above involve a form of auditorimotor disconnection, then auditorimotor PAS could be used as a novel adjuvant therapy to assist in the re/establishment of appropriate sensorimotor mappings.

ACKNOWLEDGMENTS

Paul F. Sowman is supported by the National Health and Medical Research Council, Australia (#543438, #1003760) and the

Australian Research Council (DE130100868). The authors would like to thank AC Etchell for his comments on the manuscript.

REFERENCES

- Alibiglou, L., and Mackinnon, C. D. (2012). The early release of planned movement by acoustic startle can be delayed by transcranial magnetic stimulation over the motor cortex. *J. Physiol.* 590, 919–936. doi: 10.1111/jphysiol.2011.219592
- Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer*. [Computer program] (Version 5.3.51), retrieved 2 June 2013 from <http://www.praat.org/>
- Cacace, A. T. (2003). Expanding the biological basis of tinnitus: crossmodal origins and the role of neuroplasticity. *Hear. Res.* 175, 112–132. doi: 10.1016/s0378-5955(02)00717-7
- Cai, S., Beal, D. S., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (2014). Impaired timing adjustments in response to time-varying auditory perturbation during connected speech production in persons who stutter. *Brain Lang.* 129, 24–29. doi: 10.1016/j.bandl.2014.01.002
- Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., and Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation. *PLoS One* 7:e41830. doi: 10.1371/journal.pone.0041830
- Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., and Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Hum. Brain Mapp.* 9, 183–191. doi: 10.1002/(sici)1097-0193(200004)9:4<183::aid-hbm1>3.0.co;2-z
- D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., and Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain Lang.* 118, 9–14. doi: 10.1016/j.bandl.2011.02.007
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385. doi: 10.1016/j.cub.2009.01.017
- Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402. doi: 10.1046/j.0953-816x.2001.01874.x
- Flöel, A., Ellger, T., Breitenstein, C., and Knecht, S. (2003). Language perception activates the hand motor cortex: implications for motor theories of speech perception. *Eur. J. Neurosci.* 18, 704–708. doi: 10.1046/j.1460-9568.2003.02774.x
- Forbes, A., and Sherrington, C. (1914). Acoustic reflexes in the decerebrate cat. *Amer. J. Physiol.* 35, 367–376.
- Ford, J. M., Gray, M., Faustman, W. O., Heinks, T. H., and Mathalon, D. H. (2005). Reduced gamma-band coherence to distorted feedback during speech when what you say is not what you hear. *Int. J. Psychophysiol.* 57, 143–150. doi: 10.1016/j.ijpsycho.2005.03.002
- Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/JNEUROSCI.4107-11.2012
- Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *Int. J. Lang. Commun. Disord.* 36, 149–171. doi: 10.1080/13682820118418
- Jayaram, G., and Stinear, J. W. (2008). Contralateral paired associative stimulation increases paretic lower limb motor excitability post-stroke. *Exp. Brain Res.* 185, 563–570. doi: 10.1007/s00221-007-1183-x
- Kumpulainen, S., Mrachacz-Kersting, N., Peltonen, J., Voigt, M., and Avela, J. (2012). The optimal interstimulus interval and repeatability of paired associative stimulation when the soleus muscle is targeted. *Exp. Brain Res.* 221, 241–249. doi: 10.1007/s00221-012-3165-x
- Kurt, S., Fisher, S. E., and Ehret, G. (2012). Foxp2 mutations impair auditory-motor association learning. *PLoS One* 7:e33130. doi: 10.1371/journal.pone.0033130
- Lane, H., and Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *J. Speech Lang. Hear. Res.* 14, 677–709.
- Langguth, B., Eichhammer, P., Zowe, M., Kleinjung, T., Jacob, P., Binder, H., et al. (2005). Altered motor cortex excitability in tinnitus patients: a hint at crossmodal plasticity. *Neurosci. Lett.* 380, 326–329. doi: 10.1016/j.neulet.2005.01.069
- Liotti, M., Ingham, J. C., Takai, O., Paskos, D. K., Perez, R., and Ingham, R. J. (2010). Spatiotemporal dynamics of speech sound perception in chronic developmental stuttering. *Brain Lang.* 115, 141–147. doi: 10.1016/j.bandl.2010.07.007
- Mayor-Dubois, C., Zesiger, P., Van der Linden, M., and Roulet-Perez, E. (2014). Nondeclarative learning in children with specific language impairment: predicting regularities in the visuomotor, phonological and cognitive domains. *Child Neuropsychol.* 20, 14–22. doi: 10.1080/09297049.2012.734293
- Michou, E., Mistry, S., Rothwell, J., and Hamdy, S. (2013). Priming pharyngeal motor cortex by repeated paired associative stimulation: implications for dysphagia neurorehabilitation. *Neurorehabil. Neural Repair* 27, 355–362. doi: 10.1177/1545968312469837
- Møller, A. R. (2003). Pathophysiology of tinnitus. *Otolaryngol. Clin. North Am.* 36, 249–266, v–vi. doi: 10.1016/S0030-6665(02)00170-6
- Möttönen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825. doi: 10.1523/JNEUROSCI.6018-08.2009
- Mrachacz-Kersting, N., Fong, M., Murphy, B. A., and Sinkjaer, T. (2007). Changes in excitability of the cortical projections to the human tibialis anterior after paired associative stimulation. *J. Neurophysiol.* 97, 1951–1958. doi: 10.1152/jn.01176.2006
- Mrachacz-Kersting, N., Kristensen, S. R., Niazi, I. K., and Farina, D. (2012). Precise temporal association between cortical potentials evoked by motor imagination and afference induces cortical plasticity. *J. Physiol.* 590, 1669–1682. doi: 10.1113/jphysiol.2011.222851
- Murakami, T., Restle, J., and Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia* 49, 2045–2054. doi: 10.1016/j.neuropsychologia.2011.03.034
- Murakami, T., Sakuma, K., Nomura, T., Uemura, Y., Hashimoto, I., and Nakashima, K. (2008). Changes in somatosensory-evoked potentials and high-frequency oscillations after paired-associative stimulation. *Exp. Brain Res.* 184, 339–347. doi: 10.1007/s00221-007-1103-0
- Namasivayam, A. K., and van Lieshout, P. (2008). Investigating speech motor practice and learning in people who stutter. *J. Fluency Disord.* 33, 32–51. doi: 10.1016/j.jfludis.2007.11.005
- Neef, N. E., Jung, K., Rothkegel, H., Pollok, B., von Gudenberg, A. W., Paulus, W., et al. (2011). Right-shift for non-speech motor processing in adults who stutter. *Cortex* 47, 945–954. doi: 10.1016/j.cortex.2010.06.007
- Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *J. Neurolinguistics* 25, 382–407. doi: 10.1016/j.jneuroling.2010.02.011
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso Del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U S A* 103, 7865–7870. doi: 10.1073/pnas.0509989103
- Ridding, M. C., and Flavel, S. C. (2006). Induction of plasticity in the dominant and non-dominant motor cortices of humans. *Exp. Brain Res.* 171, 551–557. doi: 10.1007/s00221-005-0309-2
- Roy, A. C., Craighero, L., Fabbri-Destro, M., and Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *J. Physiol. Paris* 102, 101–105. doi: 10.1016/j.jphysparis.2008.03.006
- Sato, M., Buccino, G., Gentilucci, M., and Cattaneo, L. (2010). On the tip of the tongue: modulation of the primary motor cortex during audiovisual speech perception. *Speech Commun.* 52, 533–541. doi: 10.1016/j.specom.2009.12.004
- Schecklmann, M., Volberg, G., Frank, G., Hadersdorfer, J., Steffens, T., Weisz, N., et al. (2011). Paired associative stimulation of the auditory system: a proof-of-principle study. *PLoS One* 6:e27088. doi: 10.1371/journal.pone.0027088
- Stefan, K., Kunesch, E., Benecke, R., Cohen, L. G., and Classen, J. (2002). Mechanisms of enhancement of human motor cortex excitability induced by interventional paired associative stimulation. *J. Physiol.* 543, 699–708. doi: 10.1113/jphysiol.2002.023317
- Stephan, K. E., Friston, K. J., and Frith, C. D. (2009). Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophr. Bull.* 35, 509–527. doi: 10.1093/schbul/sbn176
- Suppa, A., Li Voti, P., Rocchi, L., Papazachariadis, O., and Berardelli, A. (2013). Early visuomotor integration processes induce LTP/LTD-like plasticity in the human motor cortex. *Cereb. Cortex* doi: 10.1093/cercor/bht264. [Epub ahead of print].

- Thabit, M. N., Ueki, Y., Koganemaru, S., Fawi, G., Fukuyama, H., and Mima, T. (2010). Movement-related cortical stimulation can induce human motor plasticity. *J. Neurosci.* 30, 11529–11536. doi: 10.1523/JNEUROSCI.1829-10.2010
- Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054
- Uy, J., Ridding, M. C., Hillier, S., Thompson, P. D., and Miles, T. S. (2003). Does induction of plastic change in motor cortex improve leg function after stroke? *Neurology* 61, 982–984. doi: 10.1212/01.wnl.0000078809.33581.1f
- Watkins, K. E., Strafella, A. P., and Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994. doi: 10.1016/s0028-3932(02)00316-0
- Wolters, A., Schmidt, A., Schramm, A., Zeller, D., Naumann, M., Kunesch, E., et al. (2005). Timing-dependent plasticity in human primary somatosensory cortex. *J. Physiol.* 565, 1039–1052. doi: 10.1113/jphysiol.2005.084954

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 March 2014; accepted: 18 May 2014; published online: 03 June 2014.

Citation: Sowman PF, Dueholm SS, Rasmussen JH and Mrachacz-Kersting N (2014) Induction of plasticity in the human motor cortex by pairing an auditory stimulus with TMS. *Front. Hum. Neurosci.* 8:398. doi: 10.3389/fnhum.2014.00398

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Sowman, Dueholm, Rasmussen and Mrachacz-Kersting. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

or the arguments surrounding the cause of the disorder, but rather a hypothesis as to one of the possible causes of stuttering. The proposal that timing is important for speech (see Lashley, 1951; Martin, 1972; Strait et al., 2011) and even speech disorders like specific language impairment (Tallal et al., 1993) dyslexia (Goswami, 2011) or indeed stuttering (Alm, 2004, 2010) is not new. In the later case, the idea that stuttering relates to a deficit of timing follows from the observation that regular external stimulation temporarily alleviates stuttering (see for a revision, Alm, 2004; Snyder et al., 2009). The novel aspect of this article is that it expands on previous research suggesting that dysfunction within a brain network that supports internal timing [comprised of the basal ganglia (BG) and the supplementary motor area (SMA)] is causing stuttering and that a secondary system which utilizes external timing cues to sequence movements [comprised of the cerebellum (CB), the premotor cortex (PMC) and the right inferior frontal gyrus (IFG)] is compensating for stuttering. Specifically, we propose that an internal timing network (ITN), largely equivalent to the “medial system” proposed by Goldberg (1985) is involved in internally timed movement (movement performed in the absence of external timing cues) and is causally related to stuttering. We further propose that an external timing network (ETN), largely equivalent to the “lateral system” proposed by Goldberg (1985), with the addition of the right IFG, is involved in externally timed movement (movement performed in the presence of external timing cues) and provides a substrate for timing compensation in stuttering. Importantly, we are not suggesting that neural deficits in structures underlying timing is the sole cause of stuttering, but rather one of many possible deficits that could lead to stuttering. In this section, we first present multimodal neuroimaging evidence for the possible causal involvement of ITN in stuttering before moving on to discuss putative compensatory roles of the ETN.

There is ongoing debate as to whether some brain regions are specifically dedicated to processing time or whether the capacity to process time is intrinsic to each region of the brain directly through the activation of sensory processes (for review see Ivry and Schlerf, 2008). There already exist reviews outlining the cognitive and neural architecture proposed for how we represent a sense of time (e.g., Buhusi and Meck, 2005), how different sensory networks interact with core timing networks across different tasks (e.g., Merchant et al., 2013) as well as evidence for common timing mechanisms across manual and oral movements (e.g., Franz et al., 1992). While the questions of how and where time is processed in the brain are of considerable practical and theoretical interest, such a discussion is outside the scope of this article. Here we argue that the ETN is primarily active when an individual is timing their movement to an external rhythm and that it is particularly active during early exposure to rhythm or when the rhythm is difficult and is not easily internalized. In contrast to this, the ITN is primarily active when an individual is making rhythmic motor movements that are not specifically timed to an external stimulus. Importantly, the two systems can be active simultaneously such as when an individual is pacing their movements to an external stimulus and is internalizing that rhythm. Practically, this means that results of functional magnetic resonance imaging (fMRI) studies may show no difference

in brain activation between conditions that supposedly bias internally or externally-timed movements; however, disruption of these systems via inhibitory transcranial magnetic stimulation (TMS) should yield selective interference in behavioral performance. What follows is a brief overview of studies supporting a dissociation between the ITN and the ETN in timing tasks.

There is strong support for the involvement of the ITN during timing tasks from a number of fMRI, magnetoencephalography (MEG), lesion and TMS studies. For example, a recent fMRI study has found that the BG and the SMA tend to be active when movements are internally as opposed to being externally timed (Coull et al., 2013). Similarly, it has been shown using finger tapping tasks, that the BG and the SMA are active during the continuation phase (no external pacing stimulus, hence an internally-timed process) but not the synchronization phase (with external pacing, hence externally-timed) of the task (Rao et al., 1997). In particular, the BG are more active during the performance or tracking of simple rhythms, i.e., those that are easier to internalize, compared to complex rhythms (Grahn and Rowe, 2009, 2013; Geiser et al., 2012). The fact that fMRI studies show an overlap of neural activity during synchronization and continuation tapping (e.g., Jäncke et al., 2000; Jantzen et al., 2004) provides little support for a functional distinction between brain networks supporting internal and external timing; however, evidence from lesion and TMS does support such a dissociation between the ITN and the ETN and their respective functions. Studies show that individuals with bilateral lesions to the BG perform poorly on the continuation phase of the finger-tapping task (Coslett et al., 2010) and are also poor at adjusting to accelerations and decelerations in tempo (Schwartz et al., 2011). Disruption of the SMA by inhibitory TMS impairs accuracy of continuation tapping whilst leaving the accuracy of synchronization tapping intact (Halsband et al., 1993).

There is also evidence for the involvement of CB and the PMC in the ETN. Inhibitory TMS of the CB has been shown to disrupt synchronization to auditory (Del Olmo et al., 2007) and visual pacing (Theoret et al., 2001; Koch et al., 2007). This disruption appears to be selective because lesions to the CB do not affect performance during the continuation phase of the finger-tapping task (Spencer et al., 2003). Likewise, a number of studies show that inhibitory TMS of the left PMC disrupts the synchronization tapping (Pollok et al., 2008; Bijsterbosch et al., 2011) and that this effect is specific to external pacing, as no effect of TMS is observed on continuation tapping (Del Olmo et al., 2007) or when tapping in the presence of, but not in time with, a scrambled beat (Kornysheva and Schubotz, 2011). Taken together, there indeed appears to be a functional dissociation of the ITN and the ETN in healthy adults. We now turn to neuroimaging studies to demonstrate how these systems are impaired in people who stutter.

NEUROIMAGING STUDIES OF THE INTERNAL TIMING NETWORK IN PWS

A number of neuroimaging studies implicate the BG or components thereof in the etiology of stuttering. For example, when comparing the fluent and dysfluent speech of people who stutter (PWS) to people who do not stutter (PWDS), Wu et al. (1995) found that PWS exhibited less activity in the caudate during both

dysfluent speech and fluent speech. This lowered activity was suggested to be a trait marker for stuttering. The BG has also been related to the most typical symptoms of stuttering at an individual level (Jiang et al., 2012). These authors elicited stuttering during a sentence completion task and classified repetitions, pauses and prolongations as being either least typical or most typical of stuttering based on patterns of haemodynamic responses. Jiang et al. (2012) found that one of the activation patterns contributing to this separation of most and least typical symptoms was a reduction in BG activation. Although the aforementioned studies provide a correlative link between the putative ITN and stuttering, they do not unambiguously support the notion that the ITN *causes* stuttering. Because those studies were conducted mainly in adults, and stuttering is a disorder that appears in childhood, it can therefore be hard to determine whether anomalous BG activations observed in PWS are related to the cause of stuttering or are compensations for it.

In contrast, structural and functional abnormalities in children who stutter (CWS) are likely to be more indicative of the causative agents in stuttering because children have not had as much time to adapt to stuttering as adults. Chang and Zhu (2013), examined functional connectivity in CWS and children who do not stutter (CWDS) aged 3–9 and found reduced levels of connectivity between the putamen and the SMA, superior temporal gyrus (STG) and CB and similarly between the SMA and the putamen, STG and CB. Chang and Zhu (2013) concluded that CWS exhibited reduced activity in areas responsible for self-paced movement as compared to CWDS. Similarly, a recent voxel based morphometry (VBM) study conducted in CWS, found less gray matter volume in the bilateral inferior frontal gyri and the left putamen but more gray matter volume in the right rolandic operculum and the right STG relative to CWDS (Beal et al., 2013). In another study, Foundas et al. (2013) measured the volume of the caudate in right-handed boys who stutter and compared them to right-handed boys who did not stutter. They found that male CWS exhibited significantly less volume in the right caudate as compared to male CWDS. These studies suggest that even at a very young age, CWS exhibit abnormalities in structure and connectivity in the ITN. A recent MEG study examined lateralization of brain functions in preschool CWS and CWDS during a picture-naming task (Sowman et al., 2014). These authors found that speech was strongly left lateralized in both groups. Although not explicitly focusing on the ITN, this study demonstrates that much of the abnormal activation observed in the cortical right hemisphere in adults is the result of years of compensation for stuttering rather than being causally related to it. Moreover, that there were no differences between CWS and CWDS in cortical activations further hints at the possibility that stuttering is caused by deficiencies in subcortical regions. Overall, these studies provide strong support for viewing stuttering as a disorder of the BG. Since the BG seems responsible for internal timing of movement, they provide indirect support that stuttering is a disorder of internally timed movement.

To implicate the ITN in stuttering, structural or functional abnormalities should be evident in these structures in both children and adults who stutter and the neural deficit necessary to cause stuttering should be present irrespective of whether or not a subject is performing a task. Ingham et al. (2012) examined

speech during oral reading and monologs as well as during a rest condition and found that PWS were different to PWDS in both the medial (ITN) and lateral (ETN) systems proposed by Alm (2004). PWS had significantly more activity in the BG (including the left putamen) during an eyes closed rest condition but significantly less activity during speaking conditions. This was thought to result in difficulties in performing fine-grained movement that may extend to speech and explain the fact that other studies observed increased activation of these regions in speech conditions like oral reading and monolog. More specifically though, if it is the case that the BG are overactive during rest and not just underactive during speech, it would indicate abnormalities in stuttering are not solely confined to speech. That is to say, the problem spans a number of domains because there are functional differences in neural activation occurring in the absence of speech.

If abnormalities of the ITN are causally related to stuttering, then it could be expected that effective speech therapy should produce measurable changes in the neural activity of these structures rather than in the areas compensating for stuttering. To this end, Giraud et al. (2008) examined neural activity using fMRI before and after speech therapy in a group of PWS. Therapy consisted of 3 weeks of undergoing an inpatient program focusing on biofeedback for syllable prolongation, soft voice onset and smooth sound transition. The researchers found that activity in the caudate positively correlated with stuttering severity before speech therapy but not after. Since the caudate was positively correlated with severity rather than negatively correlated with it, the speech therapy appeared to target causal rather than compensatory regions.

Similarly, if the ITN is related to stuttering this will not only be reflected in measures of neural activity but also in terms of the connections within the ITN. Lu et al. (2010) used structural equation modeling to compare causal relationships and function in the ITN in PWS and PWDS during a picture-naming task. Although there were no significant differences between stuttering and non-stuttering speakers in the output of the SMA to the BG, there were significant differences between the groups in the output of the BG to the SMA. More specifically, whereas PWDS showed a strong negative projection from the BG to the pre-SMA, PWS showed a positive projection from the BG to the pre-SMA. Lu et al. (2010) interpreted their finding of abnormal output of the BG to the SMA as reflecting the difficulties PWS have in updating the timing and sequencing of movement. Interestingly, like Lu et al. (2010), a number of other studies have also shown altered patterns of activity in the SMA in relation to the perception and planning of speech in stuttering (Chang et al., 2009, 2011). Taken together, these findings, are consistent with the notion that stuttering is the result of dysfunctional processes that engage core structures within the proposed ITN: the BG and the SMA.

LESION STUDIES OF THE ITN IN PWS

If dysfunction in the ITN is thought to cause stuttering, then it follows that damage to these regions may result in stuttering. When stuttering develops following a lesion to the brain it is known as acquired or neurogenic stuttering (for review see Lundgren et al., 2010). There is evidence that damage to the ITN results in stuttering. For example a recent study by Tani and Sakai (2011) examining five patients with BG lesions (two with

bilateral putamen lesions, two patients with bilateral BG lesions and one patient with a left putamenal lesion) but without aphasia, found that they exhibited dysfluencies such as syllable repetitions, part word repetitions and frequent blocks. Importantly, these patients' symptoms mimicked the characteristics of developmental stuttering in that almost all stuttering occurred on the initial syllable of a word. In a number of case studies, Ciabarra et al. (2000) describe a right-handed woman with a left BG lesion, and a woman with a left corona radiata, putamenal and subinsular infarct who both stuttered. Similarly, a number of different case studies have reported the onset of stuttering following damage to the SMA (Alexander et al., 1987; Ackermann et al., 1996; Chung et al., 2004). Furthermore, direct electrical stimulation of the SMA has also been shown to induce stuttering (Penfield and Welch, 1951). These findings are consistent with the notion that damage to the SMA can cause speech disorders and that the SMA is linked with the rhythmic control of speech (Jonas, 1981). This and other works have prompted investigation into the role of the SMA in rhythmic movements of the mouth (MacNeilage and Davis, 2001) as well as dissociations between the pre-SMA and the SMA-proper in rhythmic timing (Schwartz et al., 2012).

NEUROIMAGING STUDIES OF THE ETN SYSTEM IN PWS

There are studies hinting that deficits to the ITN are causing stuttering, but what proof is there that the ETN is recruited to compensate for this? To answer this question, we turn to fMRI studies of PWS. Braun et al. (1997) found the CB to be overactive in PWS during stuttered and fluent speech and it has been suggested that this is a compensatory mechanism for stuttering (see also Alm, 2004). In a meta-analysis of PWS, Brown et al. (2005) identified three neural signatures of stuttering. These neural signatures were the absence of auditory activation bilaterally, the over-activation of the right IFG and the over-activation of the CB. These findings have since been partially replicated by Lu et al. (2010) who found over-activation of the right IFG and the CB (but not the absence of bilateral auditory activation) and interpreted them as compensating for stuttering. Ingham et al. (2012) examined speech during oral reading and monologs as well as rest, finding that PWS exhibited increased cerebellar activity which was negatively associated with stuttering, indicating that the ETN may indeed be compensating for the ITN. A similar study, examined resting state functional connectivity of PWS before and after speech therapy in stuttering and non-stuttering adults (Lu et al., 2012). These authors found increased resting-state-functional-connectivity between the midline CB and a network of regions (comprised of the medial frontal gyrus, the SMA and the left IFG) at rest for PWS relative to PWDS. For the PWS who received intervention as compared to the PWS who did not receive intervention (and PWDS), the resting-state-functional-connectivity in the midline CB returned to normal levels and was correlated with an increase in fluency. As such, Lu et al. (2012) suggested the CB was likely compensating in stuttering. In addition to these, other studies have associated the CB with compensatory activation in PWS (e.g., De Nil et al., 2008; Watkins et al., 2008).

While there is overlap in the neural structures responsible for external timing and compensation for stuttering, it does not automatically follow that the ETN is compensating for deficits in

internal timing in PWS. However, there is fMRI evidence showing that the CB and the right IFG specifically compensate for deficits in the BG with respect to timing tasks in those who have Parkinson's Disease (PD). For example, Jahanshahi et al. (2010), investigated the differences in neural activation between PD patients and controls in and the synchronization continuation task. They also examined the effect of administering apomorphine (a non-selective dopamine agonist) on neural activation in the PD patients. Results showed that for healthy controls synchronization and continuation tapping (relative to a control reaction time task) was associated with significantly greater activation in the nucleus accumbens and caudate, a pattern not found in PD patients. In contrast, individuals with PD showed greater activation in the bilateral cerebellar hemispheres, right thalamus and left midbrain during both phases of finger tapping. Administration of apomorphine to the PD patients appeared to normalize activity, both increasing the connectivity between the caudate and putamen and frontal regions as well as decreasing activity in the CB. Thus, the authors suggested that increased cerebellar activation was likely compensating for the impaired functioning of the BG. Sen et al. (2010) found increased cerebellar-thalamo-cortical (CTC) activation as PD progressed, perhaps indicating an increasing need to compensate for loss of function in the striato-thalamo-cortical networks (STC). This increase was only observed during continuation tapping and was not evident during synchronization tapping suggesting that the CTC (i.e., the ETN) was compensating for the STC (i.e., the ITN). The dissociation between the ITN and the ETN may seem problematic given both the CB (part of the ETN) and the SMA (part of the ITN) are thought to compensate for deficits in the BG during self initiated hand movements in the early stages of PD (Eckert et al., 2006). Nevertheless, this could suggest that part of the ITN (the SMA) may still be able to compensate for deficits in other parts of the ITN (the BG) when degeneration is not particularly severe.

COMPENSATION BY THE RIGHT IFG IN STUTTERING

An increasing number of studies have reported anomalous activation of the right IFG in a variety of speech tasks (e.g., Fox et al., 1996; Brown et al., 2005; Sowman et al., 2012) in PWS. Several studies found that increases in right IFG activation during overt reading (Preibisch et al., 2003; Lu et al., 2010) that were positively correlated with speech fluency in PWS and thought to be a non-specific compensatory mechanism because the activation was not specifically related to speech production. Examining the effect of external auditory pacing on the speech of PWS Toyomura et al. (2011) found that, relative to a PWDS, the PWS showed more activation in the right IFG (along with bilateral auditory cortices) during both choral speaking and when speaking in time with an isochronous metronome. There are also reports of increased right frontal connections in adults who began stuttering as children (i.e., developmental stuttering) relative to adults who began stuttering later in life following a psychological trigger and without evidence of brain injury (Chang et al., 2010). This evidence suggests that the longer a PWS has been compensating for their stuttering, the greater the activity in the right IFG.

It is worth noting that Goldberg's formulation of the lateral system (upon which the ETN partially maps) does not contain the

right IFG. Why then should right IFG be considered a part of an ETN that compensates for a dysfunctional ITN in stuttering? This question is particularly relevant when considering that the simplest explanation for right IFG involvement in stuttering is that it compensates for deficits in the left IFG (see Kell et al., 2009). Kell et al. (2009) associate the left IFG with processing of rhythm and sensorimotor feedback and it is possible that the right IFG may perform a similar function. Recently, the right IFG has been recognized as part of a “core timing network” (Wiener et al., 2010) that is recognized to be strongly connected both functionally and structurally to the ITN (Kung et al., 2013; Brittain and Brown, 2014). In particular, the right IFG may only become active when a task is more demanding. That is to say, the difficulty of compensating for deficits in internal timing by external timing regions might account for why there was over-activation of only the CB during speech, but not the right IFG during rest in PWS (Lu et al., 2012). A second, though not mutually exclusive explanation is that while the CB is able to compensate for timing deficits, its ability to do so is limited. This is evident in the case of individuals with PD where behavioral performance worsened despite increases in compensatory activation in the CB (Sen et al., 2010). A similarly limited ability of the cerebellar systems to compensate for deficits in timing may be occurring in PWS as evidenced by the reduced integrity of cerebellar tracts in both the left and the right hemispheres (Connally et al., 2013). Since the ETN has a limited capacity to compensate for deficits in the ITN, the assistance of the right IFG may be required to maintain normal timing functions. A third possible explanation is that the model proposed by Goldberg (1985) (where the ETN is comprised of the CB and the PMC) is incomplete and requires the addition of the right IFG as a secondary part of the system. Importantly, the right IFG is not likely to be the only region that is compensating for stuttering. There are many other regions like the orbitofrontal cortex that could be compensating depending on the task and motor regions involved (see Kell et al., 2009; Sowman et al., 2012). Our contention is that the right IFG forms part of a network that compensates for deficient internal timing.

BEHAVIORAL STUDIES OF TIMING IN PWS

If stuttering is the result of dysfunction in the ITN, and the ITN is important for timing, then it follows that PWS should exhibit deficits in behavioral performance on timing tasks. To this end several groups have found significant differences in asynchrony and variability of tapping between PWS and PWDS. For example, measuring the timing variability of reading sentences or nursery rhymes or tapping, Cooper and Allen (1977) found that PWS were consistently more variable in the length of time it took them to read sentences, paragraphs or nursery rhymes, and in their inter-tap intervals compared to PWDS. Brown et al. (1990) found that PWS were slower and less variable than PWDS at repeating the phrase “ah” and tapping their fingers as at their own pace compared to PWDS, findings they interpreted to represent less flexible timing systems which were more susceptible to breakdown. Similarly, when examining the timing intensity and variability of externally timed speech, Boutsen et al. (2000) showed that although both PWS and PWDS exhibited similar intensities when producing syllables, PWS were significantly more variable in their inter-onset vocalization times (analogous

to the inter tap interval in tapping tasks). Additionally, Zelaznik et al. (1997) found that PWS were more variable on bimanual finger tapping (something more demanding than unimanual finger tapping) relative to PWDS. Similarly, Hulstijn et al. (1992) found that on a task which required the coordination of finger tapping and vocal responses (tapping in time with vocalizing the word “pip”), PWS exhibited greater variability than PWDS. More recently, Olander et al. (2010) compared hand-clapping variability in CWS and CWDS. While there was no difference in mean clapping rate, there were significant differences between groups in the variability of the clapping rate. This variability was bimodally distributed, with 60% of CWS showing variability that was greater than the worst performing CWDS. The remaining CWS showed variability in clapping that overlapped with that of the CWDS. Interestingly, this number approximately corresponded to the number of children that spontaneously recover and whose stuttering persists. As a result, the authors suggested that the motor timing deficit may be predictive of recovery from stuttering. Later, Foundas et al. (2013) found that when male CWS were required to tap as fast as possible in a given time period, most were better when tapping with their left rather than right hands as compared to most male CWDS who showed an advantage for their right hand. A recent behavioral study has found robust differences in tapping performance between CWS who stutter compared to CWDS (Falk et al., 2014). In contrast to the CWDS, the CWS not only tapped earlier and were less consistent in tapping, but also failed to improve with age.

However, a number of studies have compared the asynchrony and variability of PWS and PWDS on externally or internally timed vocal or oral motor movements and found similar levels of variance between the groups (e.g., Hulstijn et al., 1992; Melvine et al., 1995). Similar results have been obtained by Zelaznik et al. (1994) who compared PWS and PWDS on externally and internally timed manual responses for isochronous intervals and found that the groups did not differ in behavioral performance. Likewise, Max and Yudman (2003) found PWS and PWDS displayed highly similar levels of asynchrony and variability for finger tapping and producing vocalizations for multiple isochronous intervals. Overall, the behavioral studies investigating the timing abilities of PWS have produced mixed results. While some studies have found differences between PWS and PWDS, many have failed to find differences between groups. From this research, it might seem appropriate to conclude that stuttering is not a disorder of timing and that the links between stuttering and deficits in production of timed limb movements is tenuous at best. One possible explanation is that motor control of limbs and speech is different both centrally and peripherally (Kent, 2000). However if this were indeed the case, then it would be hard to explain why some studies did find significant differences between PWS and PWDS in non-speech motor tasks. Moreover, there is evidence of common timing systems across modalities (Franz et al., 1992) and it has been stressed that the behavioral differences between PWS and PWDS are not confined to the speech production system and instead appear to be generalized deficits (Max et al., 2003). There are other possible explanations for the failure to find behavioral differences between groups which can, in part, be attributed to compensatory neural activity and task difficulty.

TENTATIVE SUGGESTIONS FOR TIMING DEFICITS IN PWS

The substantial number of studies finding no difference in timing behavior in PWS and PWDS is inconsistent with the notion that stuttering could be considered a disorder of timing. How then can we resolve these seemingly paradoxical findings with the consistent observation that neural regions involved in internal timing display anomalous function and structure in stuttering? The absence of a difference at a behavioral level does not imply the absence of differences at a *neural* level. Even a task as simple as tapping a finger or vocalizing to a metronome recruits a complex network of brain regions each with a variety of different functions (Repp and Su, 2013). Moreover, there may be differences at the neural level in the absence of differences at the behavioral level precisely because PWS are compensating for deficits in internal timing. Such a possibility is highlighted by the findings of Neef et al. (2011), who, utilizing inhibitory TMS, showed PWS did not exhibit behavioral differences in timing prior to stimulation but did exhibit behavioral differences subsequent to stimulation. If the suggestion that PWS demonstrate similar behavioral performance as a result of re-organization is plausible, then PWS should exhibit compensatory neural activity in regions associated with external timing of movement that are specifically compensating for deficits in the internal timing of movements. This indeed appears to be the case as both the CB and the right IFG seem to be compensatory regions in stuttering; both appear to be associated with timing, and both may specifically be compensating for deficits in the BG's control of timing tasks. Although speculative, this strongly suggests that the compensatory response to stuttering that occurs during speech is occurring as a result of deficits in the ITN. It perhaps explains why, in some studies at least, PWS have not shown differences in asynchrony (the difference in time between taps and the pacing signal) or variability (in the time between taps) on tapping tasks compared to PWDS. However, any failure to find a difference between these groups may also be attributed to task related effects such as the motoric or temporal complexity.

Many of the behavioral studies investigating timing abilities in PWS employed simple motoric and temporal tasks. Tapping at isochronous intervals is, as a task, relatively easy and this ease may explain a lack of differences in behavioral performance between PWS and PWDS, a problem that may extend to differences in regional brain activation in neuroimaging studies. Imaging data from early research on finger movements shows that the amount of cerebral blood flow to a particular region depends upon the complexity of the task (Shibasaki et al., 1993). Simple tasks are, *ipso facto*, not sufficiently motorically demanding to engage parts of the brain normally employed in more complex tapping tasks and which are impaired in PWS. This principle has been demonstrated experimentally in a number of studies. For example, Zelaznik et al. (1994) failed to find behavioral differences when comparing unimanual tapping performance, but successfully found differences in the same group of stuttering participants when examining bimanual tapping at an isochronous interval (Zelaznik et al., 1997). Similarly, increasing the syntactic complexity of words surrounding a to-be-repeated phrase, decreased speech motor stability for PWS as compared to PWDS (Kleinow and Smith, 2000).

In the same way that increasing the difficulty of the motor movement associated with the task could better reveal differences (should they exist) in behavioral performance and neural activation, so too could placing more strain on the systems governing temporal control of movements. Whereas Webster (1985) failed to find a difference in behavioral performance for PWS during bimanual tapping in a 1:1 ratio (that is one tap of the right hand for every tap of the left hand), Webster (1990) found that PWS took a substantially longer time to tap the required number of times when tapping in a ratio of 2:1 (that is two taps of the left hand for each tap of the right hand) than PWDS. Tapping at an uneven ratio (2:1) places significantly more demand on the neural systems governing timing than does tapping in an even ratio (1:1). This suggests that PWS are much less efficient in coordinating motor output to complex temporal patterns. Similarly, Lewis et al. (2004) demonstrated that parametrically increasing the number of different intervals in a series of tones resulted in a corresponding increase in neural activation in regions associated with timing. These studies show that, increasing the demands on temporal processing is more likely to yield differences in behavior and by extension, in neural activation. This is particularly relevant in the case of speech since speech is rarely perfectly isochronous but rather quasi-periodic (Martin, 1972). Speech contains multiple levels of temporal complexity (Kotz and Schwartz, 2010; Goswami and Leong, 2013) and is therefore substantially more demanding than tapping at an isochronous interval or in a 1:1 ratio. That is to say, differences in the complexity of rhythms required for speech and finger tapping may explain why most timed movements are relatively normal in PWS. Additionally, the timing required for speech control is robust to interference so difficulties in timing movements or speech may only become evident under increased cognitive loads (e.g., Saltuklaroglu et al., 2009). If PWS were compared to PWDS on a tapping task that contained a similar degree of temporal complexity usually required by speech, then clinically meaningful differences in behavior are likely to emerge. While there is a theoretical distinction between motor and temporal complexity, in practice, this distinction may not be so clear. Using near infrared spectroscopy (a means to measure the level of deoxygenated blood from the scalp somewhat analogous to how fMRI measures neural activity) Koenraadt et al. (2013) found that the two may not be mutually exclusive. Tapping at multiple frequencies activated larger portions of the motor cortex than tapping at single frequencies. The extent to which manipulating motoric and temporal complexity are able to elicit behavioral differences in timing between PWS and PWDS remains to be tested by future research. Yet, even if these tasks are unable to elicit such differences in PWS, future research investigating the overlap between stuttering and timing should consider the use of neuroimaging techniques.

DIRECTIONS FOR FUTURE RESEARCH

There appears to be a vast gap in the stuttering literature particularly with respect to neuroimaging and brain stimulation of timing tasks. In particular, we know of no fMRI or positron emission tomography (PET) studies that specifically examined internally or externally timed movements in PWS using either simple or complex temporal intervals despite the long theoretical

history of an association between deficient timing and stuttering. The timing deficits we propose to exist in PWS are only tentative suggestions and remain to be verified by future research. Our proposal can nevertheless be used to generate a number of testable hypotheses. For example, it could be hypothesized that PWS show impaired behavioral performance and corresponding neural activation in tasks that require the internal timing of movements (the continuation phase of a finger tapping task) as opposed to the external timing of movements (the synchronization phase of a finger tapping task).

Likewise to the best of our knowledge, there are no studies investigating neural oscillations in PWS in response to isochronous or non-isochronous tones either by passive listening, finger tapping or vocalizations. Given the role of neural oscillations in timing (Arnal, 2012), it would be interesting to investigate how they might differ between PWS and PWDS in the context of a timing task. With respect to studies of brain stimulation, no studies have yet examined the effect of disruptive TMS on the right IFG, the SMA or the CB in PWS in a timing task. Although speculative, it might be expected that tapping in time to a metronome (external timing) will be relatively unimpaired because PWS can rely on the CB and premotor cortices much in the same way as non-stuttering adults do. However for self-paced tapping it might be expected that following inhibitory TMS to the right IFG, PWS will be significantly impaired because they cannot rely on either the right IFG or the BG. In contrast, PWDS will be able to rely on the BG, but not the right IFG. The compensatory function of the right IFG in stuttering is biologically plausible in that it forms part of a core timing-network (Wiener et al., 2010), is functionally interconnected with the BG (Kung et al., 2013) and is utilized for the processing of speech rhythm (Geiser et al., 2012).

While this article focused on the neural correlates of the ITN and the ETN during the perception and production of rhythmic movements and stimuli, there are many other tasks that probe these networks. The finger-tapping task is a continuous task that is often conducted in the presence of a regular external stimulus. It is possible that the regular external stimulus reduces behavioral variability and (possibly the associated) neural activity much in the same way that it is able to temporarily induce fluency in PWS. It would therefore be prudent to examine the timing abilities of PWS on tasks that do not contain such regular stimuli or where there is a disruption to the external stimuli. In line with the hypothesis of impaired internal timing and the hypothesized compensatory increases in regions associated with the processing of external timing of movements, it might be expected that PWS are more reliant on external cues. As such it would be interesting to test abilities of PWS to judge whether a “test interval” is longer or shorter than a “reference interval” and how these judgments are influenced by the presence of a “distractor interval” that they must ignore (see Rao et al., 2001). To this end, we know of no studies that have examined temporal judgment deficits in PWS either behaviorally or neurologically. More generally, if it is demonstrated that PWS exhibit deficits in timing, it would be particularly interesting to see if there is any dissociation between these different types of timing tasks or modalities; There may for example, be a dissociation between motor timing or judgment duration or between auditory and visual timing.

CONCLUDING REMARKS

In conclusion, we provide a theoretical framework with which to view stuttering as a disorder of timing. This paper reviews converging evidence from neuroimaging and brain stimulation experiments showing a great degree of overlap between the structures engaged in the internal timing of movements and the regions thought to be causally involved in stuttering. We also provide evidence of overlap between the neural structures engaged in the external timing of movement and link them with compensatory activity in PWS. We further highlight significant gaps in the literature and suggest avenues for further research motivated by this overarching theory. More generally, this article highlights anomalies in the functional activations and the structural anatomy of the areas involved in the processing of time in stuttering, that are linked to the dysfluent production of speech and should motivate further research in the field.

ACKNOWLEDGMENTS

We thank Paul Tawadros for his valuable comments on the manuscript. This work was supported by the Australian Research Council (DE130100868).

REFERENCES

- Ackermann, H., Hertrich, I., Zieglers, W., Bitzer, M., and Bien, S. (1996). Acquired dysfluencies following infarction of the left mesiofrontal cortex. *Aphasiology* 10, 409–417. doi: 10.1080/02687039608248420
- Alexander, M., Naeser, M., and Palumbo, C. (1987). Correlations of subcortical CT lesion sites and aphasia profiles. *Brain* 110, 961–991. doi: 10.1093/brain/110.4.961
- Alm, P. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. *J. Commun. Disord.* 37, 325–369. doi: 10.1016/j.jcomdis.2004.03.001
- Alm, P. A. (2010). “The dual premotor model of cluttering and stuttering: a neurological framework,” in *Proceedings of the First World Conference on Cluttering*, eds K. Bakker, L. Raphael, and F. Myers (Katarino), 200.
- Ambrose, N. G. (2004). Theoretical perspectives on the cause of stuttering. *Contemp. Issues Commun. Sci. Disord.* 31, 80–91.
- Arnal, L. H. (2012). Predicting “When” using the motor system’s beta-band oscillations. *Front. Hum. Neurosci.* 6:225. doi: 10.3389/fnhum.2012.00225
- Beal, D. S., Gracco, V. L., Brettschneider, J., Kroll, R. M., and De Nil, L. F. (2013). A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. *Cortex* 49, 2151–2161. doi: 10.1016/j.cortex.2012.08.013
- Bijsterbosch, J. D., Lee, K. H., Dyson-Sutton, W., Barker, A. T., and Woodruff, P. W. (2011). Continuous theta burst stimulation over the left pre-motor cortex affects sensorimotor timing accuracy and supraliminal error correction. *Brain Res.* 1410, 101–111. doi: 10.1016/j.brainres.2011.06.062
- Boutsen, F. R., Brutton, G. J., and Watts, C. R. (2000). Timing and intensity variability in the metronomic speech of stuttering and nonstuttering speakers. *J. Speech Lang. Hear. Res.* 43, 513–520. doi: 10.1044/jslhr.4302.513
- Braun, A. R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J. M., et al. (1997). Altered patterns of cerebral activity during speech and language production in developmental stuttering. An H2 (15) O positron emission tomography study. *Brain* 120, 761–784. doi: 10.1093/brain/120.5.761
- Brittain, J. S., and Brown, P. (2014). Oscillations and the basal ganglia: motor control and beyond. *Neuroimage* 85, 637–647. doi: 10.1016/j.neuroimage.2013.05.084
- Brown, C. J., Zimmermann, G. N., Linville, R. N., and Hegmann, J. P. (1990). Variations in self-paced behaviors in stutterers and nonstutterers. *J. Speech Lang. Hear. Res.* 33, 317–323.
- Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., and Fox, P. T. (2005). Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain Mapp.* 25, 105–117. doi: 10.1002/hbm.20140
- Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* 6, 755–765. doi: 10.1038/nrn1764

- Chang, S. E., Horwitz, B., Ostuni, J., Reynolds, R., and Ludlow, C. L. (2011). Evidence of left inferior frontal–premotor structural and functional connectivity deficits in adults who stutter. *Cereb. Cortex* 21, 2507–2518. doi: 10.1093/cercor/bhr028
- Chang, S. E., Kenney, M. K., Loucks, T. M., and Ludlow, C. L. (2009). Brain activation abnormalities during speech and non-speech in stuttering speakers. *Neuroimage* 46, 201–212. doi: 10.1016/j.neuroimage.2009.01.066
- Chang, S. E., Synnæstvedt, A., Ostuni, J., and Ludlow, C. L. (2010). Similarities in speech and white matter characteristics in idiopathic developmental stuttering and adult-onset stuttering. *J. Neurolinguistics* 23, 455–469. doi: 10.1016/j.jneuroling.2008.11.004
- Chang, S. E., and Zhu, D. C. (2013). Neural network connectivity differences in children who stutter. *Brain* 136, 3709–3726. doi: 10.1093/brain/awt275
- Chung, S. J., Im, J. H., Lee, J. H., and Lee, M. C. (2004). Stuttering and gait disturbance after supplementary motor area seizure. *Mov. Disord.* 19, 1106–1109. doi: 10.1002/mds.20136
- Ciabarra, A. M., Elkind, M. S., Roberts, J. K., and Marshall, R. S. (2000). Subcortical infarction resulting in acquired stuttering. *J. Neurol. Neurosurg. Psychiatry* 69, 546–549. doi: 10.1136/jnnp.69.4.546
- Connally, E., Ward, D., Howell, P., and Watkins, K. E. (2013). Disrupted white matter in language and motor tracts in developmental stuttering. *Brain Lang.* 131, 25–36. doi: 10.1016/j.bandl.2013.05.013
- Couture, E. G., and Kelly, E. M. (1991). Young stutterers' nonspeech behaviors during stuttering. *J. Speech Lang. Hear. Res.* 34, 1041–1056.
- Cooper, M. H., and Allen, G. D. (1977). Timing control accuracy in normal speakers and stutterers. *J. Speech Lang. Hear. Res.* 20, 55.
- Coslett, H. B., Wiener, M., and Chatterjee, A. (2010). Dissociable neural systems for interval timing: evidence from subjects with basal ganglia lesions. *PLoS ONE* 5:e10324. doi: 10.1371/journal.pone.0010324
- Coull, T. C., Davranche, K., Nazarian, B., and Vidal, F. (2013). Functional anatomy of timing differs for production versus prediction of time intervals. *Neuropsychologia* 51, 309–319. doi: 10.1016/j.neuropsychologia.2012.08.017
- Craig, A., Blumgart, E., and Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. *J. Fluency Disord.* 34, 61–71. doi: 10.1016/j.jfludis.2009.05.002
- Del Olmo, M. F., Cheeran, B., Koch, G., and Rothwell, J. C. (2007). Role of the cerebellum in externally paced rhythmic finger movements. *J. Neurophysiol.* 98, 145–152. doi: 10.1152/jn.01088.2006
- De Nil, L. F., Beal, D. S., Lafaille, S. J., Kroll, R. M., Crawley, A. P., and Gracco, V. L. (2008). The effects of simulated stuttering and prolonged speech on the neural activation patterns of stuttering and nonstuttering adults. *Brain Lang.* 107, 114–123. doi: 10.1016/j.bandl.2008.07.003
- Eckert, T., Peschel, T., Heinze, H. J., and Rotte, M. (2006). Increased pre-SMA activation in early PD patients during simple self-initiated hand movements. *J. Neurol.* 253, 199–207. doi: 10.1007/s00415-005-0956-z
- Falk, S., Müller, T., and Bella, S. D. (2014). Sensorimotor synchronization in stuttering children and adolescents. *Proc. Soc. Behav. Sci.* 126, 206–207. doi: 10.1016/j.sbspro.2014.02.375
- Foundas, A. L., Cindass, R. Jr., Mock, J. R., and Corey, D. M. (2013). Atypical caudate anatomy in children who stutter 1, 2. *Percept. Mot. Skills* 116, 528–543. doi: 10.2466/15.10.PMS.116.2.528-543
- Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., et al. (1996). A PET study of the neural systems of stuttering. *Nature* 382, 158–161. doi: 10.1038/382158a0
- Franz, E. A., Zelaznik, H. N., and Smith, A. (1992). Evidence of common timing processes in the control of manual, orofacial, and speech movements. *J. Motor Behav.* 24, 281–187. doi: 10.1080/00222895.1992.9941623
- Geiser, E., Notter, M., and Gabrieli, J. D. (2012). A corticostriatal neural system enhances auditory perception through temporal context processing. *J. Neurosci.* 32, 6177–6182. doi: 10.1523/JNEUROSCI.5153-11.2012
- Giraud, A. L., Neumann, K., Bachoud-Levi, A. C., von Gudenberg, A. W., Euler, H. A., Lanfermann, H., et al. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain Lang.* 104, 190–199. doi: 10.1016/j.bandl.2007.04.005
- Goldberg, G. (1985). Supplementary motor area structure and function. *Behav. Brain Sci.* 8, 567–616. doi: 10.1017/S0140525X00045167
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* 15, 3–10. doi: 10.1016/j.tics.2010.10.001
- Goswami, U., and Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives? *Lab. Phonol.* 4, 67–92. doi: 10.1515/lp-2013-0004
- Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/JNEUROSCI.2018-08.2009
- Grahn, J. A., and Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cereb. Cortex* 23, 913–921. doi: 10.1093/cercor/bhs083
- Halsband, U., Ito, N., Tanji, J., and Freund, H.-J. (1993). The role of premotor cortex and the supplementary motor area in the temporal control of movement in man. *Brain* 116, 243–266. doi: 10.1093/brain/116.1.243
- Hulstijn, W., Summers, J. J., van Lieshout, P. H., and Peters, H. F. (1992). Timing in finger tapping and speech: a comparison between stutters and fluent speakers. *Hum. Mov. Sci.* 11, 113–124. doi: 10.1016/0167-9457(92)90054-F
- Ingham, R. J., Grafton, S. T., Bothe, A. K., and Ingham, J. C. (2012). Brain activity in adults who stutter: similarities across speaking tasks and correlations with stuttering frequency and speaking rate. *Brain Lang.* 122, 11–24. doi: 10.1016/j.bandl.2012.04.002
- Iverach, L., O'Brian, S., Jones, M., Block, S., Lincoln, M., Harrison, E., et al. (2009). Prevalence of anxiety disorders among adults seeking speech therapy for stuttering. *J. Anxiety Disord.* 23, 928–934. doi: 10.1016/j.janxdis.2009.06.003
- Ivry, R. B., and Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends Cogn. Sci.* 12, 273–280. doi: 10.1016/j.tics.2008.04.002
- Jahanshahi, M., Jones, C. R., Zijlmans, J., Katzenschlager, R., Lee, L., Quinn, N., et al. (2010). Dopaminergic modulation of striato-frontal connectivity during motor timing in Parkinson's disease. *Brain* 133, 727–745. doi: 10.1093/brain/awq012
- Jäncke, L., Loose, R., Lutz, K., Specht, K., and Shah, N. J. (2000). Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. *Cogn. Brain Res.* 10, 51–66. doi: 10.1016/S0926-6410(00)00022-7
- Jantzen, K. J., Steinberg, F. L., and Kelso, J. A. (2004). Brain networks underlying human timing behavior are influenced by prior context. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6815–6820. doi: 10.1073/pnas.0401300101
- Jiang, J., Lu, C., Peng, D., Zhu, C., and Howell, P. (2012). Classification of types of stuttering symptoms based on brain activity. *PLoS ONE* 7:e39747. doi: 10.1371/journal.pone.0039747
- Jonas, S. (1981). The supplementary motor region and speech emission. *J. Commun. Disord.* 14, 349–373. doi: 10.1016/0021-9924(81)90019-8
- Kell, C. A., Neumann, K., von Kriegstein, K., Posenenske, C., von Gudenberg, A. W., Euler, H., et al. (2009). How the brain repairs stuttering. *Brain* 132, 2747–2760. doi: 10.1093/brain/awp185
- Kent, R. D. (2000). Research on speech motor control and its disorders: a review and prospective. *J. Commun. Disord.* 33, 391–428. doi: 10.1016/S0021-9924(00)00023-X
- Kleinow, J., and Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *J. Speech Lang. Hear. Res.* 43, 548–559. doi: 10.1044/jslhr.4302.548
- Koch, G., Oliveri, M., Torriero, S., Salerno, S., Lo Gerfo, E., and Caltagirone, C. (2007). Repetitive TMS of cerebellum interferes with millisecond time processing. *Exp. Brain Res.* 179, 291–299. doi: 10.1007/s00221-006-0791-1
- Koenraadt, K. L., Duysens, J., Meddeler, B. M., and Keijsers, N. L. (2013). Hand tapping at mixed frequencies requires more motor cortex activity compared to single frequencies: an fNIRS study. *Exp. Brain Res.* 231, 231–237. doi: 10.1007/s00221-013-3686-y
- Kornysheva, K., and Schubotz, R. I. (2011). Impairment of auditory–motor timing and compensatory reorganization after ventral premotor cortex stimulation. *PLoS ONE* 6:e21421. doi: 10.1371/journal.pone.0021421
- Kotz, S. A., and Schwartz, M. (2010). Cortical speech processing unplugged: a timely subcortical-cortical framework. *Trends Cogn. Sci.* 14, 392–399. doi: 10.1016/j.tics.2010.06.005
- Kung, S. J., Chen, J. L., Zatorre, R. J., and Penhune, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* 25, 401–420. doi: 10.1162/jocn_a_00325
- Langevin, M., Packman, A., and Onslow, M. (2010). Parent perceptions of the impact of stuttering on their preschoolers and themselves. *J. Commun. Disord.* 43, 407–423. doi: 10.1016/j.jcomdis.2010.05.003
- Lashley, K. S. (1951). “The problem of serial order in behavior,” in *Cerebral Mechanisms in Behaviour*, ed L. A. Jeffress (New York, NY: Wiley), 112–136.

- Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., and Miall, R. C. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialization, synchronization, and continuation phases of paced finger tapping. *Neuropsychologia* 42, 1301–1312. doi: 10.1016/j.neuropsychologia.2004.03.001
- Lu, C., Chen, C., Peng, D., You, W., Zhang, X., Ding, G., et al. (2012). Neural anomaly and reorganization in speakers who stutter: A short-term intervention study. *Neurology* 79, 625–632. doi: 10.1212/WNL.0b013e31826356d2
- Lu, C., Peng, D., Chen, C., Ning, N., Ding, G., Li, K., et al. (2010). Altered effective connectivity and anomalous anatomy in the basal ganglia-thalamocortical circuit of stuttering speakers. *Cortex* 46, 49–67. doi: 10.1016/j.cortex.2009.02.017
- Lundgren, K., Helm-Estabrooks, N., and Klein, R. (2010). Stuttering following acquired brain damage: a review of the literature. *J. Neurol.* 23, 447–454. doi: 10.1016/j.jneurology.2009.08.008
- MacNeilage, P. F., and Davis, B. L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. *Curr. Opin. Neurobiol.* 11, 696–700. doi: 10.1016/S0959-4388(01)00271-9
- Martin, J. (1972). Rhythmic (Hierarchical) versus serial structure in speech and other behavior. *Psychol. Rev.* 79, 487–509. doi: 10.1037/h0033467
- Max, L., Caruso, A. J., and Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *J. Speech Lang. Hear. Res.* 46, 215–232. doi: 10.1044/1092-4388(2003)017
- Max, L., and Yudman, E. A. (2003). Accuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. *J. Speech Lang. Hear. Res.* 46, 146–163. doi: 10.1044/1092-4388(2003)012
- Melvine, C., Williams, H., Bishop, J., McClenaghan, B., Cooper, W., and McDade, H. (1995). “Vocal and manual timing control of adult stutterers and nonstutterers,” in *Stuttering: Proceedings of the first World Congress on Fluency Disorders*, eds C. W. Starkweather and U. F. M. Peters (Nijmegen: University of Nijmegen), 35–38.
- Merchant, H., Harrington, D. L., and Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annu. Rev. Neurosci.* 36, 313–336. doi: 10.1146/annurev-neuro-062012-170349
- Neef, N. E., Jung, K., Rothkegel, H., Pollok, B., Von Gudenberg, A. W., Paulus, W., et al. (2011). Right-shift for non-speech motor processing in adults who stutter. *Cortex* 47, 945–954. doi: 10.1016/j.cortex.2010.06.007
- Olander, L., Smith, A., and Zelaznik, H. N. (2010). Evidence that a motor timing deficit is a factor in the development of stuttering. *J. Speech Lang. Hear. Res.* 53, 876–886. doi: 10.1044/1092-4388(2009)09-0007
- Packman, A. (2012). Theory and therapy in stuttering: a complex relationship. *J. Fluency Disord.* 37, 225–233. doi: 10.1016/j.jfludis.2012.05.004
- Penfield, W., and Welch, K. (1951). The supplementary motor area of the cerebral cortex: a clinical and experimental study. *Arch. Neurol. Psychiatry* 66, 89–317. doi: 10.1001/archneurpsyc.1951.02320090038004
- Pollok, B., Rothkegel, H., Schnitzler, A., Paulus, W., and Lang, N. (2008). The effect of rTMS over left and right dorsolateral premotor cortex on movement timing of either hand. *Eur. J. Neurosci.* 27, 757–764. doi: 10.1111/j.1460-9568.2008.06044.x
- Preibisch, C., Neumann, K., Raab, P., Euler, H. A., von Gudenberg, A. W., Lanfermann, H., et al. (2003). Evidence for compensation for stuttering by the right frontal operculum. *Neuroimage* 20, 1356–1364. doi: 10.1016/S1053-8119(03)00376-8
- Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W., and Binder, J. R. (1997). Distributed neural systems underlying the timing of movements. *J. Neurosci.* 17, 5528–5535.
- Rao, S. M., Mayer, A. R., and Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nat. Neurosci.* 4, 317–323. doi: 10.1038/85191
- Repp, B. H., and Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2
- Riva-Posse, P., Busto-Marolt, L., Schteinschnaider, Á., Martínez-Echenique, L., Cammarota, Á., and Merello, M. (2008). Phenomenology of abnormal movements in stuttering. *Parkinsonism Relat. Disord.* 14, 415–419. doi: 10.1016/j.parkreldis.2007.11.006
- Saltuklaroglu, T., Teulings, H. L., and Robbins, M. (2009). Differential levels of speech and manual dysfluency in adults who stutter during simultaneous drawing and speaking tasks. *Hum. Mov. Sci.* 28, 643–654. doi: 10.1016/j.humov.2008.08.003
- Schwartz, M., Keller, P. E., Patel, A. D., and Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behav. Brain Res.* 216, 685–691. doi: 10.1016/j.bbr.2010.09.015
- Schwartz, M., Rothermich, K., and Kotz, S. A. (2012). Functional dissociation of pre-SMA and SMA-proper in temporal processing. *Neuroimage* 60, 290–298. doi: 10.1016/j.neuroimage.2011.11.089
- Sen, S., Kawaguchi, A., Truong, Y., Lewis, M. M., and Huang, X. (2010). Dynamic changes in cerebellothalamo-cortical motor circuitry during progression of Parkinson's disease. *Neuroscience* 166, 712–719. doi: 10.1016/j.neuroscience.2009.12.036
- Shibasaki, H., Sadato, N., Lyshkow, H., Yonekura, Y., Honda, M., Nagamine, T., et al. (1993). Both primary motor cortex and supplementary motor area play an important role in complex finger movement. *Brain* 116, 1387–1398. doi: 10.1093/brain/116.6.1387
- Snyder, G. J., Hough, M. S., Blanchet, P., Ivy, L. J., and Waddell, D. (2009). The effects of self-generated synchronous and asynchronous visual speech feedback on overt stuttering frequency. *J. Commun. Disord.* 42, 235–244. doi: 10.1016/j.jcomdis.2009.02.002
- Sowman, P. F., Crain, S., Harrison, E., and Johnson, B. W. (2012). Reduced activation of left orbitofrontal cortex precedes blocked vocalization: a magnetoencephalographic study. *J. Fluency Disord.* 37, 359–365. doi: 10.1016/j.jfludis.2012.05.001
- Sowman, P. F., Crain, S., Harrison, E., and Johnson, B. W. (2014). Lateralization of brain activation in fluent and non-fluent preschool children: a magnetoencephalographic study of picture-naming. *Front. Hum. Neurosci.* 8:354. doi: 10.3389/fnhum.2014.00354
- Spencer, R. M., Zelaznik, H. N., Diedrichsen, J., and Ivry, R. B. (2003). Disrupted timing of discontinuous but not continuous movements by cerebellar lesions. *Science* 300, 1437–1439. doi: 10.1126/science.1083661
- Strait, D. L., Hornickel, J., and Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behav. Brain. Funct.* 7:44. doi: 10.1186/1744-9081-7-44
- Tallal, P., Miller, S., and Fitch, R. H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. *Ann. N.Y. Acad. Sci.* 682, 27–47. doi: 10.1111/j.1749-6632.1993.tb22957.x
- Tani, T., and Sakai, Y. (2011). Analysis of five cases with neurogenic stuttering following brain injury in the basal ganglia. *J. Fluency Disord.* 36, 1–16. doi: 10.1016/j.jfludis.2010.12.002
- Theoret, H., Haque, J., and Pascual-Leone, A. (2001). Increased variability of paced finger tapping accuracy following repetitive magnetic stimulation of the cerebellum in humans. *Neurosci. Lett.* 306, 29–32. doi: 10.1016/S0304-3940(01)01860-2
- Toyomura, A., Fujii, T., and Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *Neuroimage* 57, 1507–1516. doi: 10.1016/j.neuroimage.2011.05.039
- Watkins, K. E., Smith, S. M., Davis, S., and Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain* 131, 50–59. doi: 10.1093/brain/awn241
- Webster, W. G. (1985). Neuropsychological models of stuttering—I. Representation of sequential response mechanisms. *Neuropsychologia* 23, 263–267. doi: 10.1016/0028-3932(85)90110-1
- Webster, W. G. (1990). Evidence in bimanual finger-tapping of an attentional component to stuttering. *Behav. Brain Res.* 37, 93–100. doi: 10.1016/0166-4328(90)90084-R
- Wiener, M., Turkeltaub, P., and Coslett, H. B. (2010). The image of time: a voxel-wise meta-analysis. *Neuroimage* 49, 1728–1740. doi: 10.1016/j.neuroimage.2009.09.064
- World Health Organisation. (2010). *International Classification of Diseases (ICD)*. Geneva: World Health Organisation. Available online at: <http://apps.who.int/classifications/icd10/browse/2010/en/#F98.5>
- Wu, J. C., Maguire, G., Riley, G., Fallon, J., LaCasse, L., Chin, S., et al. (1995). A positron emission tomography [18F] deoxyglucose study of developmental stuttering. *Neuroreport* 6, 501–505. doi: 10.1097/00001756-199502000-00024
- Yairi, E., and Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *J. Fluency Disord.* 38, 66–87. doi: 10.1016/j.jfludis.2012.11.002

- Zelaznik, H. N., Smith, A., Franz, E., and Ho, M. (1997). Differences in bimanual coordination associated with stuttering. *Acta Psychol.* 96, 229–243. doi: 10.1016/S0001-6918(97)00014-0
- Zelaznik, H. N., Smith, A., and Franz, E. A. (1994). Motor performance of stutterers and nonstutterers on timing and force control tasks. *J. Motor. Behav.* 26, 340–347. doi: 10.1080/00222895.1994.9941690

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 March 2014; accepted: 08 June 2014; published online: 25 June 2014.

Citation: Etchell AC, Johnson BW and Sowman PF (2014) Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: a hypothesis and theory. *Front. Hum. Neurosci.* 8:467. doi: 10.3389/fnhum.2014.00467
This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Etchell, Johnson and Sowman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Sequencing at the syllabic and supra-syllabic levels during speech perception: an fMRI study

Isabelle Deschamps^{1,2*} and Pascale Tremblay^{1,2}

¹ Département de Réadaptation, Université Laval, Québec City, QC, Canada

² Centre de recherche de l'Institut universitaire en santé mentale de Québec, Québec City, QC, Canada

Edited by:

Patti Adank, University College
London, UK

Reviewed by:

Jonathan E. Peelle, Washington
University in St. Louis, USA
Carolyn McGettigan, Royal Holloway
University of London, UK
Samuel Evans, University College
London, UK

*Correspondence:

Isabelle Deschamps, Centre de
recherche de l'Institut universitaire
en santé mentale de Québec, 2601
rue de la Canadière, Office
F-2424A, Québec City, QC G1J 2G3,
Canada
e-mail: isabelle.deschamps.1@
ulaval.ca

The processing of fluent speech involves complex computational steps that begin with the segmentation of the continuous flow of speech sounds into syllables and words. One question that naturally arises pertains to the type of syllabic information that speech processes act upon. Here, we used functional magnetic resonance imaging to profile regions, using a combination of whole-brain and exploratory anatomical region-of-interest (ROI) approaches, that were sensitive to syllabic information during speech perception by parametrically manipulating syllabic complexity along two dimensions: (1) individual syllable complexity, and (2) sequence complexity (supra-syllabic). We manipulated the complexity of the syllable by using the simplest syllable template—a consonant and vowel (CV)—and inserting an additional consonant to create a complex onset (CCV). The supra-syllabic complexity was manipulated by creating sequences composed of the same syllable repeated six times (e.g., /pa-pa-pa-pa-pa-pa/) and sequences of three different syllables each repeated twice (e.g., /pa-ta-ka-pa-ta-ka/). This parametrical design allowed us to identify brain regions sensitive to (1) syllabic complexity independent of supra-syllabic complexity, (2) supra-syllabic complexity independent of syllabic complexity and, (3) both syllabic and supra-syllabic complexity. High-resolution scans were acquired for 15 healthy adults. An exploratory anatomical ROI analysis of the supratemporal plane (STP) identified bilateral regions within the anterior two-third of the planum temporale, the primary auditory cortices as well as the anterior two-third of the superior temporal gyrus that showed different patterns of sensitivity to syllabic and supra-syllabic information. These findings demonstrate that during passive listening of syllable sequences, sublexical information is processed automatically, and sensitivity to syllabic and supra-syllabic information is localized almost exclusively within the STP.

Keywords: syllabic information, supra-syllabic information, supratemporal plane, speech processing, language

INTRODUCTION

The speech signal is undoubtedly one of the most complex auditory signals that humans are exposed to, requiring multiple computational steps to parse and convert acoustic waves into discrete linguistic units from which meaning can be extracted. Unsurprisingly, given such complexity, the manner in which the human brain accomplishes the complex computational steps leading to the comprehension of speech remains far from understood.

Functional neuroimaging studies of speech perception offer converging evidence suggesting that the supratemporal plane (STP), and superior temporal sulcus (STS) play a critical role in the processing of speech sounds. This finding is quite robust having been observed under different types of speech perception tasks (i.e., passive listening, monitoring and discrimination tasks as well as neural adaptation paradigms) and with different types of speech stimuli (words, pseudo-words, syllables, phonemes). For instance, neuroimaging studies contrasting the neural activity evoked by speech stimuli to the neural activity associated with the processing of acoustically complex non-speech sounds or silence have reliably reported clusters of activation within the STP and/or

STS (Zatorre et al., 1992; Binder et al., 1996, 1997; Dhankhar et al., 1997; Celsis et al., 1999; Burton et al., 2000; Scott et al., 2000; Benson et al., 2001; Vouloumanos et al., 2001; Joanisse and Gati, 2003; Wilson et al., 2004; Liebenthal et al., 2005; Rimol et al., 2005; Wilson and Iacoboni, 2006; Obleser et al., 2007; Okada et al., 2010; Zhang et al., 2011; Tremblay et al., 2012). In addition, neuropsychological evidence demonstrate that bilateral lesions to the superior temporal lobes can result in pure word deafness, a deficit associated with impaired word comprehension but relatively intact ability to process non-speech sounds (Buchman et al., 1986; Tanaka et al., 1987; Poeppel, 1996). While both functional and neuropsychological studies provide strong evidence regarding the importance of the STP and STS for the perception of speech sounds, the specific contribution of each of the sub-regions that form this large cortical area to speech perception is still uncertain; whether it is related to the processing of acoustical, sublexical, or lexical information.

Several neuroimaging studies have contrasted the neural activity evoked by different sublexical units (e.g., consonant clusters, phonemes, syllables) to the processing of non-speech or unintelligible speech sounds (e.g., sinewave analogs, tones,

environmental sounds, noise, spectrally rotated syllables, silence) to isolate speech specific processes. These studies reported reliable activation within supratemporal regions [e.g., the superior temporal gyrus (STG), the transverse temporal gyrus (TTG), and planum temporale (PT)], the STS, the middle temporal gyrus (MTG) and, in some instances, in the inferior parietal lobule (IPL), and the inferior frontal gyrus (IFG) (Demonet et al., 1992; Zatorre et al., 1992; Binder et al., 1994; Dhankhar et al., 1997; Giraud and Price, 2001; Vouloumanos et al., 2001; Liebenthal et al., 2005; Rimol et al., 2005; Pulvermüller et al., 2006; Obleser et al., 2007; Tremblay et al., 2012). The consistency of the STP and STS results in studies using words or sublexical units suggest that these regions might be involved in the conversion of acoustical information into phonological representations. However, because these studies have contrasted different types of sublexical units to non-speech or unintelligible speech sounds, the level of processing (e.g., acoustical/phonetic, phonemic, syllabic, supra-syllabic) at which mechanisms implemented within the STP and STS operate remains unclear.

Neuroimaging studies in which phonological mechanisms are engaged by the use of an explicit task (discrimination, rhyming) can more readily target specific mechanisms operating at different sublexical levels (phonemic, syllabic, supra-syllabic) and offer valuable insights into the functional contribution of STP regions to the perception of speech sounds. For instance, STP and STS activation have been reported in studies using a variety of auditory tasks: phonetic discrimination (Burton et al., 2000), rhyming (Booth et al., 2002), syllable identification (Liebenthal et al., 2013), monitoring (Rimol et al., 2005), and phonemic judgments (Jacquemot et al., 2003). Other studies using a neural adaptation paradigm to target phonetic processing have also identified regions within the STP that responded more strongly to stimuli that were part of different phonemic categories than those that fell within the same phonemic category (Dehaene-Lambertz et al., 2005; Joannisse et al., 2007). Taken together, these studies support the notion of a key involvement of the STP and STS in processing sounds at different levels (phonemic, syllabic). However, despite their importance, studies using explicit speech perception tasks requiring judgments on speech sounds probably recruit to greater extent phonological processes than do more naturalistic speech tasks. It is therefore unclear whether similar regions would be recruited in the absence of a task. It is also unclear whether phonological mechanisms operating at different levels (phonemic, syllabic, supra-syllabic) engage the same or different neural networks. Despite the scarcity of studies addressing this issue, in a recent functional magnetic resonance imaging (fMRI) study, McGettigan et al. (2011) manipulated both the complexity of syllabic and supra-syllabic information in pseudowords during a passive listening task. Syllabic complexity was manipulated by varying the number of consonant clusters (0 vs. 2) and supra-syllabic complexity was manipulated by varying the number of syllables (2 vs. 4). An effect of supra-syllabic complexity was observed in the bilateral PT. However, no positive¹

effect of syllabic complexity was reported. In contrast, Tremblay and Small (2011), also using fMRI, varied syllabic complexity as indexed by the presence or absence of consonant clusters during the passive listening of words and found that the right PT was sensitive to the syllabic complexity manipulation, supporting the idea that the supratemporal cortex plays a role in processing syllabic information (Grabski et al., 2013).

One question that arises from this literature is whether specific sublexical processes can be localized to specific regions within the STP and STS. In the current experiment, we were interested in investigating the distinct and shared effects of syllabic and supra-syllabic complexity on brain activity during the processing of auditory sequences. To this aim, we parametrically manipulated phonological complexity along two dimensions (1) individual syllable complexity (presence or absence of a consonant cluster in the syllable onset) and (2) sequence-level complexity (the ordering of syllables within a sequence). Given the importance of the STP and STS in the processing of auditory information, we conducted an exploratory anatomical ROI analysis focusing on a fine-grain parcellation of the supratemporal cortex and STS based on our previous work (Tremblay et al., 2012, 2013) to determine whether sub-regions within the STP and STS process similar or different kind of sublexical information during passive speech perception (i.e., syllabic or supra-syllabic). In these prior studies, we demonstrated that sub-regions within the STP exhibited different patterns of sensitivity to speech sounds during speech perception and production, suggesting that the STP contains a mosaic of functionally distinct areas. It is therefore possible that sub-regions within the STP are processing the speech signal in different manners and at different levels, with some focusing on spectral information, while others on syllable- or sequence-level information. Based on the results from our previous studies, we hypothesized that some sub-regions within the STP (in particular the PT) and STS would show similar patterns of activation for both manipulations while others would show a preference for one manipulation. For example, we expected the primary auditory cortex to be sensitive to both manipulations, as both syllabic and supra-syllabic complexity increase acoustic complexity. We also expected the PT to be sensitive to the syllabic manipulation based on previous results (Tremblay and Small, 2011).

MATERIALS AND METHODS

PARTICIPANTS

The participants were 15 healthy right-handed (Oldfield, 1971) native French speakers (9 females; 26.8 ± 4.8 years; range 21–34, education 17.3 ± 1.9 years), with normal hearing and no history of language or neurological/neuropsychological disorders. Hearing was assessed using pure tone audiometry (clinical audiometer, AC40, Interacoustic) for each ear separately for the following frequencies: 0.25, 0.5, 1, 2, 3, 4, 5, 8, 12, and 16 kHz. Then for each participant, a standard pure tone average (PTA: average of threshold at 0.5, 1, and 2 kHz) was computed for the left (17.13 ± 3.78 dB) and right ear (18.68 ± 3.17 dB), since most of the speech sounds fall within this range (Stach, 2010). All participants were screened for depression (Yesavage et al., 1982) and their cognitive functioning was evaluated using the Montreal Cognitive Assessment scale (MOCA) (Nasreddine et al., 2005).

¹ The authors reported several brain areas in which blood-oxygen-level dependent (BOLD) signal magnitude was higher for pseudowords without consonant clusters than for pseudowords containing consonant clusters.

All participants were within normal range on the MOCA (i.e., 26/30 or better) and none of the participants were depressive. The study was approved by the committee on research ethics of the Institut Universitaire en santé mentale de Québec (#280-2012).

STIMULI AND TASK

The experimental task consisted in listening passively (i.e., without performing a task) to sequences of syllables. To investigate sublexical phonological processing, we used sequences of syllables instead of pseudowords to avoid lexical effects. Prior research has demonstrated that pseudowords, given their close resemblance to words, activate regions involved in lexical access and in some cases they do so to an even greater extent than words (Newman and Twieg, 2001; Burton et al., 2005). Thus, the use of pseudowords renders the dissociation between lexical and sublexical phonological processing extremely difficult. For this reason, we decided to use syllable strings rather than words to alleviate potential lexical effects. The degree of complexity of each sequence was manipulated along two phonological dimensions: syllabic and supra-syllabic complexity. Each factor had two levels (simple or complex), resulting in a 2×2 experimental design matrix (See Table 1).

Syllabic complexity refers to the presence or absence of a consonant cluster (e.g., /gr/): simple syllables were composed of a single consonant and vowel (CV) and complex syllables were composed of a consonant cluster and a vowel (CCV). Supra-syllabic complexity refers to the number of different syllables in a sequence: simple sequences were composed of the same syllable repeated six times (e.g., /ba-ba-ba-ba-ba-ba/) and complex sequences were composed of three different syllables each repeated twice (e.g., /ba-da-ga-ba-da-ga/). While these two manipulations increase phonological complexity, they target different levels of processing: syllabic (individual unit) and supra-syllabic (sequence of units).

All syllables were created by selecting among five frequent French vowels, which included two front vowels (/i/, /ε/), two back vowels (/o/, /u/), and one central vowel (ə), and 12 frequent French consonants, which included four labial consonants (/b/, /p/, /v/, /f/), four coronal consonants (/d/, /n/, /t/, /l/) and four dorsal consonants (/g/, /ʒ/, /k/, /ʁ/). These vowels and consonants were combined to form 60 simple syllables (CV) and 60 complex syllables (CCV). Each syllable was repeated a total of three times (i.e., in three different sequences). Six-syllable sequences were created by producing sequences of three different syllables twice (/pa-ta-ka-pa-ta-ka/), or by repeating one syllable six times (/pa-pa-pa-pa-pa-pa/). A native young adult male French speaker from Québec City pronounced all syllable sequences naturally in a sound attenuated booth. Each sequence was recorded five times

and the best exemplar was selected to use in the experiment. The syllable sequences were recorded at 44.1 KH using a unidirectional microphone connected to a sound card (Fast Track C-400, M-audio), saved directly to disk using Sound Studio 4.5.4 (Felt Tip Software, NY, USA), and edited offline using Wave Pad Sound Editor 4.53 (NHC Software, Canberra, Australia). Each syllable sequence was edited to have an average duration of 2400 ms. The duration of the syllable sequences was the same across all experimental conditions (i.e., 2400 ms). The root mean square (RMS) intensity was then normalized across all sound files. Individual sequences were not repeated during the course of the experiment.

PROCEDURE

This experimental paradigm resulted in four conditions of 30 trials each, for a total of 120 trials. Each trial lasted 6.5 s. A resting baseline condition was interleaved with the experimental conditions (60 trials). The randomization of the experimental and baseline conditions was optimized using Optseq2 (<http://surfer.nmr.mgh.harvard.edu/optseq/>). The four conditions were equally divided into two runs. A passive listening experimental paradigm was used; participants were not required to produce any overt response. All stimuli were presented during the delay in acquisition (see Image acquisition section below) using Presentation Software (Neurobehavioral System, CA, USA) through high-quality MRI-compatible stereo electrostatic earplugs (Nordic Neurolab, Norway), which provide 30 dB of sound attenuation.

IMAGE ACQUISITION

A 3 T Philips Achieva TX MRI scanner was used to acquire anatomical and functional data for each participant. Structural MR images were acquired with a T1-weighted MPRAGE sequence ($TR/TE = 8.2/3.7$ ms, flip angle = 8° , isotropic voxel size = 1 mm^3 , 256×256 matrix, 180 slices/volume, no gap). Single-shot EPI BOLD functional images were acquired using parallel imaging, with a SENSE reduction factor of 2 to reduce the number of phase encoding steps and speed up acquisition. In order to ensure that syllables were intelligible, a sparse image acquisition technique (Eden et al., 1999; Edmister et al., 1999; Hall et al., 1999; Gracco et al., 2005) was used. A silent period of 4360 ms was interleaved between each volume acquisition. The syllable sequences were presented 360 ms after the onset of the silent period. One hundred and eighty functional volumes were acquired across 2 runs ($TR/TE = 6500/30$ ms, volume acquisition = 2140 ms; delay in TR 4360 ms, 40 axial slices parallel to AC/PC, voxel size = $3 \times 3 \times 3$, no gap; matrix = 80×80 ; FoV = 240×240 mm). This study was part of a larger project, which also included a speech production task and a speech perception in noise task². Those two tasks will not be discussed as part of this manuscript. The speech perception task that is the focus of the present manuscript was always presented first to participants, followed by the speech production task and the speech perception

Table 1 | Experimental conditions.

Code	Syllable type	Sequence type	Examples
SS	Simple	Simple	/ba-ba-ba-ba-ba-ba/
SC	Simple	Complex	/fo-de-ro-fo-de-ro/
CS	Complex	Simple	/kli-kli-kli-kli-kli-kli/
CC	Complex	Complex	/bri-dre-klou-bri-dre-klou/

²Not all participants took part in all three tasks. Here we report the data from 15 young adults, whereas Bilodeau-Mercure et al. (2014) reported the data for a subset (11) of these participants, who performed the speech perception in noise task.

in noise task. Participants were not told until the production task that they would be required to produce speech. This was done in order to avoid priming subvocal rehearsal during the speech perception task. The speech perception in noise task has been reported elsewhere (Bilodeau-Mercure et al., 2014).

DATA ANALYSIS

fMRI time-series analyses

All functional time-series were motion-corrected, time-shifted, de-spiked and mean-normalized using AFNI (version 10.7, intel 64; Cox, 1996). All time points that occurred during excessive motion (i.e., > 1 mm) (Johnstone et al., 2006) were censored. The anatomical scan of each participant was aligned to their registered EPI time series using local Pearson correlations (Saad et al., 2009). The alignment was verified and manually adjusted when necessary. For each participant and for each run a finite impulse response ordinary least squares model was used to fit each time point of the hemodynamic response function for each of the four experimental conditions using AFNI's tent basis function (SS, SC, CS, CC). Additional regressors for the mean, the linear and quadratic trend components as well as the six motion parameters were also included. This model-free deconvolution method allows the shape of the hemodynamic response to vary for each condition rather than assuming a single response profile for all conditions (Meltzer et al., 2008). The interval modeled covered the entire volume acquisition (2.14 s), starting with stimulus onset and continuing at intervals of 6.5 s (i.e., silent period and volume acquisition) for 13 s (i.e., 2 TR). All analyses (whole-brain and ROIs) focused on the first interval (i.e., the first TR). The resulting time-series were projected onto the 2-dimensional surfaces where all subsequent processing took place.

For each participant, FreeSurfer was used to create a surface representation of the participant's MRI (Dale et al., 1999; Fischl et al., 1999) by inflating each hemisphere of the anatomical volumes to a surface representation and aligning it to a template of average curvature. SUMA was used to import the surface representations into the AFNI 3D space and to project the pre-processed time-series from the 3-dimensional volumes onto the 2-dimensional surfaces. Both the surface representations and the pre-processed time-series were standardized to a common mesh reference system (Saad et al., 2004). The time-series were smoothed on the surface to achieve a target smoothing value of 6 mm using a Gaussian full width half maximum (FWHM) filter. Smoothing on the surface as opposed to the volume ensures that white matter values are not included, and that functional data located in anatomically distant locations on the cortical surface are not averaged across sulci (Argall et al., 2006).

Group-level node-wise analyses

Whole-brain group analyses were performed using SUMA on the participants' beta values resulting from the first level analysis (Saad et al., 2004). The group level analyses focused on (1) the effect of passive auditory sequence perception on the Blood oxygenation level dependent (BOLD) signal (2) the effect of syllabic and supra-syllabic complexity on the BOLD signal during auditory sequence perception, (3) the contrast between the effect of syllabic and supra-syllabic complexity, and (4) the conjunction of the syllabic and supra-syllabic complexity effects.

To identify regions recruited during the perception of auditory sequences, a node-wise linear regression was conducted (perception > 0, one sample *t*-test option in the AFNI 3dttest++ program). To investigate the effect of syllabic and supra-syllabic complexity, a two-way repeated measure ANOVA (rANOVA) was conducted (AFNI's 3dANOVA program) with syllabic complexity (simple, complex) and supra-syllabic complexity (simple, complex) as within-subjects factors. To identify regions that exhibited a stronger response to one of the manipulations (i.e., syllabic or supra-syllabic), we computed, at the individual subject level, the effect of syllabic complexity (complex syllables - simple syllables) and the effect of supra-syllabic complexity (complex sequences - simple sequences). At the group level, the resulting *t*-maps were submitted to a paired sample *t*-test, to determine whether the two contrasts (i.e., syllable and sequence contrast) differed (AFNI 3dttest++ program). For the conjunction, we computed a map of the joint activation, for each subject, for syllabic and supra-syllabic complexity (syllabic \cap supra-syllabic). Only voxels that were significant at $p = 0.05$ (uncorrected) in both individual maps were included in the conjunction map. A group-level average of the conjunction maps was then generated. All resulting group maps were corrected for multiple comparisons using the Monte Carlo procedure implemented in FreeSurfer. This correction implements a cluster-size threshold procedure to protect against Type I error. For the first three analysis, based on the simulation results, it was determined that a family-wise error (FWE) rate of $p < 0.001$ is achieved with a minimum cluster size of 157 contiguous surface nodes, each significant at $p < 0.01$. For the conjunction analysis, we adopted a more lenient correction (a FWE rate of $p 0.05$ was achieved with a minimum cluster size of 202 contiguous surface nodes, each significant at $p < 0.05$).

Exploratory anatomical ROI analysis

To examine the role of supratemporal regions in the processing of syllabic and supra-syllabic information, we conducted an exploratory anatomical ROI analysis focusing on a set of 16 a priori selected anatomical regions. In a previous study, using a similar fine-grain parcellation, we demonstrated that several STP regions exhibited differential sensitivity pattern to auditory categories (i.e., syllables or bird songs) and sequence regularity (Tremblay et al., 2012). Here we used a similar parcellation scheme with the addition of the STS to investigate the sensitivity of these regions to syllabic and supra-syllabic information. These bilateral ROIs included the planum polare (PP), the STG, the STS, the TTG, the transverse temporal sulcus (TTS), the PT, the caudal segment of the Sylvia fissure (SF). These ROIs were anatomically defined on the participant's individual cortical surface representation using an automated parcellation scheme (Fischl et al., 2004; Desikan et al., 2006). This parcellation scheme relies on a probabilistic labeling algorithm based on the well-established anatomical convention of Duvernoy (1991). The anatomical accuracy of this method is high, approaching that of manual parcellation (Fischl et al., 2002, 2004; Desikan et al., 2006). The advantage of using anatomical (as opposed to functional) ROIs based on individual micro-anatomical landmarks is that it can capture inter-subject anatomical variability, something that is lost when using normalized templates (i.e., functional ROIs based on group level data or cytoarchitectonic maps). It is also more anatomically

precise. Thus, given that we were specifically interested in exploring the functional anatomy of the STP/STS, the choice of an anatomical ROI approach was logical.

To augment the spatial resolution of the FreeSurfer anatomical parcellation, we manually subdivided the initial parcellation of each participant's inflated surface in the following manner: the STS, the STG, the PT were subdivided into equal thirds whereas the SF, the TTG, and the TTS were subdivided into equal halves, resulting in 16 ROIs (refer to **Figure 1** and **Table 2** for details). The use of this modified FreeSurfer parcellation scheme is advantageous for several reasons: (1) it is based on a well-recognized anatomical parcellation scheme, (2) it is systematic, (3) it is easily replicable across participants and studies, and (4) it has been shown to reveal functional subdivisions within the STP (Tremblay et al., 2012).

For each participant, we extracted the mean percentage of BOLD signal change in each of the 16 resulting bilateral ROIs. First, we determined which ROIs were significantly active during the auditory perception of the sequences by testing the following hypothesis using FDR-corrected *t*-tests (Benjamini and Hochberg, 1995; Genovese et al., 2002) ($q = 0.05$): (i) perception > 0 , ($n = 32$, one-sample *t*-tests).

For each ROI that was significantly active, we conducted a three-way ANOVA with repeated measurements on the magnitude of the BOLD signal as a function of hemisphere, syllabic complexity, and supra-syllabic complexity. Within each ROI, all main effects as well as two-way and three-way interactions were examined using Bonferroni corrected paired-sample *t*-tests ($\alpha = 0.05$). Adjusted *p*-values are reported.

RESULTS

WHOLE BRAIN RESULTS

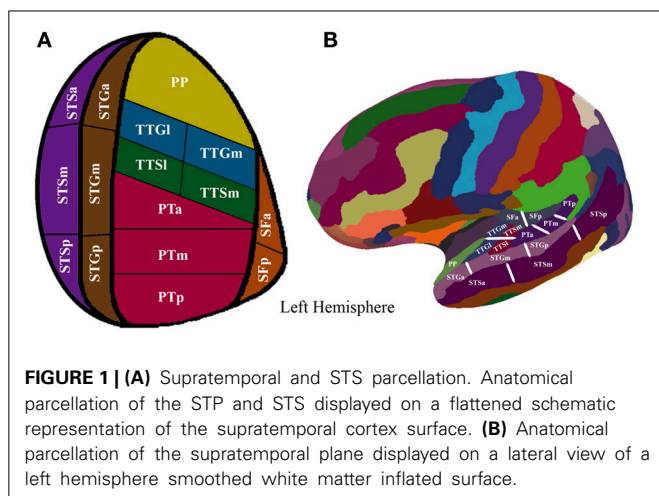
The first whole-brain analysis focused on identifying brain regions that were significantly recruited during the perception of auditory sequences regardless of syllabic and supra-syllabic complexity. The node-wise linear regression identified regions within the bilateral precentral gyrus, IFG, medial superior frontal gyrus and supratemporal cortex, as well as the left cingulate gyrus and right superior frontal gyrus that were more active than during the

perception of auditory sequences than the baseline (i.e., rest) (for details, refer to **Figure 2** and **Table 3**).

The second analysis sought to identify brain regions that were sensitive to syllabic complexity, supra-syllabic complexity. The node-wise rANOVA showed significant main effects of syllabic complexity and supra-syllabic complexity within the STP (for details, refer to **Table 4** and **Figures 3A,B**). As illustrated in **Figure 3A**, for the syllabic complexity manipulation, significant clusters of activation were observed within the left TTGI extending posteriorly into the SFp, and medially into the inferior sulcus of the insula as well as the right TTGI extending posteriorly into the SFa, laterally into the STGm and medially into the inferior circular sulcus of the insula (for details, refer to **Table 4A**). These two regions were significantly more active for the complex syllables than the simple syllables. As illustrated in **Figure 3B**, an effect of supra-syllabic complexity was found within the left STGm extending medially into the STSm, and TTSl as well as the right STGa/m, the right central sulcus and the right superior frontal gyrus. Only the clusters within the STP were significantly more active for the complex sequences (see

Table 2 | Surface description of the ROI parcellation.

Regions	Description
Superior temporal sulcus (STS)	The FreeSurfer STS ROI is bounded anteriorly by the temporal pole, medially by the STG, laterally by the MTG, and posteriorly by the IPL. We divided this region into roughly equal thirds along the rostro-caudal axis (STGp, STGm, STGa)
Superior temporal gyrus (STG)	The FreeSurfer STG ROI runs from the rostral edge of the STS to the supramarginal gyrus. It is bounded medially by the SF. We divided this region into roughly equal thirds along the rostro-caudal axis (STGp, STGm, STGp)
Planum temporale (PT)	The FreeSurfer PT ROI is bounded anteriorly by the TTS, medially by the SF, laterally by the STG, and posteriorly by the supramarginal gyrus. We divided this region into roughly equal thirds along the rostro-caudal axis (PTp, PTm, PTa)
Transverse temporal sulcus (TTS)	The FreeSurfer TTS ROI is bounded posteriorly by the PT and anteriorly by the TTG. We divided this region into two halves along the medial-lateral axis
Transverse temporal gyrus (TTG)	The FreeSurfer TTG ROI is bounded rostrally by the rostral extent of the TTS, caudally by the caudal portion of the insular cortex, laterally by the STG and medially by the SF. We divided this region in roughly equal halves along a medial-lateral axis
Caudal segment of the Sylvian Fissure (SF)	The FreeSurfer posterior SF ROI runs from the lower end of the central sulcus to the end of the posterior ascending ramus (Dahl et al., 2006). We divided this region in roughly equal halves
Planum polare (PP)	Unedited version of FreeSurfer. It is bounded rostrally by the temporal pole, caudally by the TTG, and medially by the parahippocampal gyrus



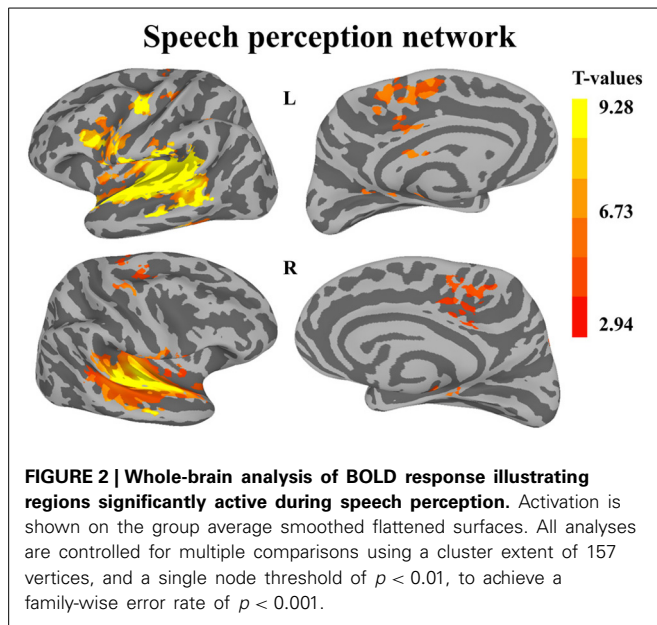


Table 4B). No significant two-way interaction between syllabic complexity and supra-syllabic complexity was found.

The third analysis sought to determine whether brain regions responded more to one complexity manipulation than the other. The node-wise t -test showed that the effect of supra-syllabic complexity was stronger than the effect of syllabic complexity within STP regions in the left STSp, STGp, and STGa, whereas the effect of syllabic complexity was stronger than the effect of supra-syllabic complexity in the left TTGL (for details, refer to **Table 5** and **Figure 3C**).

The last analysis focused on identifying regions that were sensitive to both experimental manipulations. As illustrated in **Figure 3D**, the conjunction between the syllabic complexity contrast and the supra-syllabic contrast revealed overlapping activation for both experimental manipulation within left STP regions (TTSm, TTSL, PTa, STGm), the cuneus as well as right STP regions (TTSm, TTSL, SFp), the right supramarginal gyrus, and the right subparietal sulcus. For each area that responded to both manipulations, we quantified the number of participants for which the two effects overlapped. As can be seen in **Figure 3D**, less than five participants shared common overlapping regions.

Table 3 | FWE-corrected whole-brain for the speech perception network.

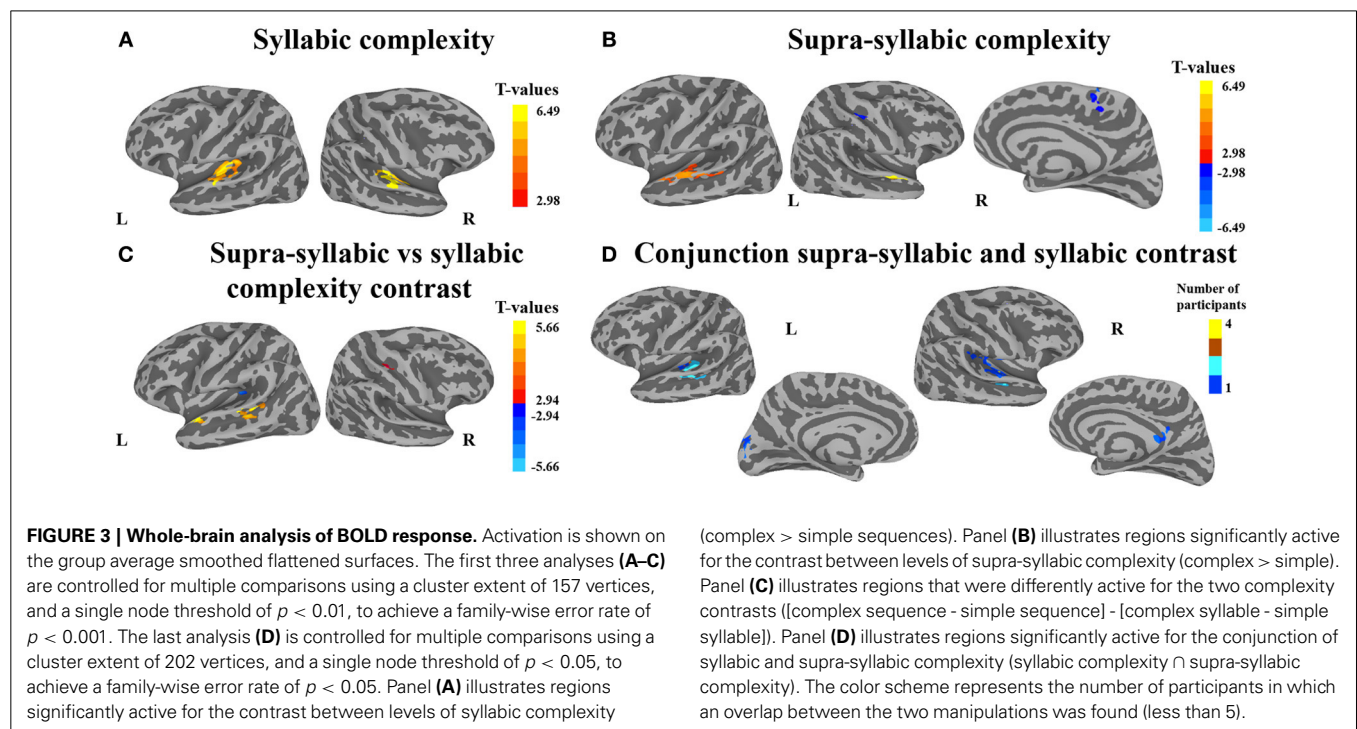
Anatomical location	Hemi	x	y	z	F-value	p-value	Cluster size (nodes)	Area (mm)
AUDITORY SEQUENCES > REST								
STGa extending into the MTG, STGp, SMG and circular sulcus of the insula (multiple clusters)	Left	-54	3	-7	9.29	$p < 0.00001$	14,323	4808
Precentral gyrus extending into the central sulcus and the inferior frontal gyrus		-58	2	19	5.91	$p < 0.00001$	2300	855
Medial superior frontal gyrus		-7	-4	56	7.29	$p < 0.00001$	1589	439
Precentral gyrus		-51	-6	45	5.74	0.00004	863	252
Cingulate gyrus and sulcus		-4	-12	38	4.59	0.00035	543	155
Central sulcus		-36	-27	48	4.29	0.00064	442	149
Lateral-occipito-temporal sulcus		-43	-50	-10	5.31	0.00009	397	121
Parieto-occipital sulcus		-15	-58	14	3.84	0.002	191	53
Supramarginal gyrus		-50	-44	47	4.74	0.00026	180	41
Cingulate gyrus		-12	-41	1	3.52	0.0031	158	24
STGm extending into the TTGL, STGa, MTG, and STSp	Right	61	-22	2	11.28	$p < 0.00001$	11,839	4200
Cingulate gyrus and sulcus (multiple clusters)		15	-23	45	6.94	$p < 0.00001$	791	177
Subcentral gyrus and sulcus		58	-4	11	4.67	0.0003	660	171
Medial superior frontal gyrus (multiple clusters)		8	-24	53	4.38	0.0005	524	143
Central sulcus		45	-9	38	7.74	$p < 0.00001$	270	102
Inferior circular sulcus of the insula and PP (two clusters)		42	0	-20	4.62	0.0003	363	100
Middle frontal gyrus and precentral sulcus (two clusters)		41	1	47	5.34	0.00008	359	86
Superior circular sulcus of the insula		38	-16	22	5.29	0.00009	171	61
Central sulcus		31	-28	50	3.87	0.0015	196	51
Superior frontal sulcus		25	0	47	4.33	0.00059	176	50
Precentral sulcus		19	-9	62	6.10	0.00002	204	47
Superior temporal sulcus		48	-54	21	-6.21	0.00002	297	41
Superior parietal gyrus		13	-74	44	4.94	0.0002	205	40
Parahippocampal gyrus		19	-31	-8	4.32	0.0006	201	39

All coordinates are in MNI space and represent the peak surface node for each of the cluster (FWE: $p = 0.001$, minimum cluster size: 157 contiguous surface nodes, each significant at $p < 0.01$). When more than one activation foci is listed, this means that the cluster had multiple peaks or was not continuous.

Table 4 | FWE-corrected whole-brain BOLD results.

Anatomical location	Hemi	x	y	z	t-value	p-value	Cluster size (nodes)	Area (mm)
(A) SYLLABIC COMPLEXITY								
TTGI extending into the SFp and the inferior circular sulcus of the insula	Left	-55	-14	2	5.84	0.00005	1618	596
TTGI extending into the SFa, STGm, and inferior circular sulcus of the insula	Right	63	-14	4	5.26	0.0001	1106	370
(B) SUPRA-SYLLABIC COMPLEXITY								
STGm extending into the TTSL and STSm (multiple clusters)	Left	-60	-12	-5	6.49	0.00002	1277	443
STGm and STGa	Right	61	-4	-4	5.91	0.00004	240	83
Central sulcus		37	-19	42	-4.71	0.0004	266	79
Superior frontal gyrus (multiple clusters)		8	-26	54	-6.82	0.000008	249	65

All coordinates are in MNI space and represent the peak surface node for each of the cluster (FWE: $p = 0.001$, minimum cluster size: 157 contiguous surface nodes, each significant at $p < 0.01$). T-values are reported instead of F-values. T-values were obtained by contrasting the two levels of complexity for each experimental factor while collapsing across the other one. When more than one activation foci is listed, this means that the cluster had multiple peaks or was not continuous.

**Table 5 | FWE-corrected whole-brain BOLD results.**

Anatomical location	Hemi	x	y	z	t-value	p-value	Cluster size (nodes)	Area (mm)
SUPRA-SYLLABIC COMPLEXITY > SYLLABIC COMPLEXITY								
STSp and STGp (two clusters)	Left	-50	-44	0	5.66	0.00005	600	196
STGa		-56	-2	-7	4.93	0.0002	172	78
TTGI		-37	-38	15	-4.63	0.0003	172	58
Central sulcus	Right	1	-11	16	-4.25	0.0007	211	60

All coordinates are in MNI space and represent the peak surface node for each of the cluster (FWE: $p = 0.001$, minimum cluster size: 157 contiguous surface nodes, each significant at $p < 0.01$). Two clusters or multiple clusters indicate that the activation cluster is not continuous.

EXPLORATORY SUPRA-TEMPORAL ROI ANALYSES

Only the ROIs that were significantly activated for speech perception were included in the subsequent analyses. Of the 32 ROIs, only the bilateral STSp was not significantly activated. For each remaining ROI ($n = 15$), we investigated the main effects of hemisphere, syllabic complexity, supra-syllabic complexity as well as the two-way interactions between hemisphere * syllabic complexity, hemisphere * supra-syllabic complexity, syllabic complexity * supra-syllabic complexity and three-way interaction between hemisphere * syllabic complexity * supra-syllabic complexity. Bonferroni adjusted p -values are reported.

As shown in **Figure 4**, a main effect of syllabic complexity was observed in the TTGl [$F_{(1, 14)} = 26.44$, $p = 0.0002$], the TTGm [$F_{(1, 14)} = 31.11$, $p = 0.00007$], the TTSl [$F_{(1, 14)} = 29.4$, $p = 0.00009$], the TTSm [$F_{(1, 14)} = 17.13$, $p = 0.001$], the STGm [$F_{(1, 14)} = 8.71$, $p = 0.011$], the SFp [$F_{(1, 14)} = 5.90$, $p = 0.029$], the SFa [$F_{(1, 14)} = 9.84$, $p = 0.007$], the PTa [$F_{(1, 14)} = 13.61$, $p = 0.002$] and the PTm [$F_{(1, 14)} = 4.84$, $p = 0.045$]. We then determined the type of stimuli driving the effect. For all nine regions, a stronger effect was observed for complex than simple syllables (paired sample t -tests, Bonferroni corrected). For the SFa, a significant hemisphere * syllabic complexity interaction was also observed [$F_{(1, 14)} = 8.39$, $p = 0.012$]. Paired sample t -tests revealed that the source of the interaction was due to the presence of an effect of syllabic complexity for the left SFa ($t = 4.39$, $p = 0.003$) but not the right SFa ($t = 1.358$, $p = 0.59$) (for details, refer to **Figure 3**). For the PTm, a significant syllabic

complexity * supra-syllabic complexity interaction was noted. Paired sample t -tests revealed that this interaction was due to the presence of an effect of syllabic complexity for the complex ($t = 2.95$, $p = 0.044$) but not the simple sequences ($t = 0.01$, $p = 1$) (for details, refer to **Figure 4**). For the TTSm, a significant syllabic complexity * hemisphere interaction was observed. Paired sample t -tests revealed that this interaction was due to a marginally significant difference when we computed a differential complexity score per hemisphere [complex - simple syllable] and compared these scores across hemispheres ($t = -2.51$, $p = 0.06$). A significant three-way interaction was observed in the STSa. To investigate the source of the three-way interaction, two-way interactions were computed. A two-way interaction between syllabic complexity and hemisphere was found for complex sequences [$F_{(1, 14)} = 7.32$, $p = 0.018$] but not for simple sequences [$F_{(1, 14)} = 0.413$, $p = 0.531$]. Paired sample t -tests were computed. A marginally significant difference ($t = -2.67$, $p = 0.054$) was found when we computed a differential complexity score per hemisphere [complex - simple syllable] and compared these scores across hemispheres.

The overall pattern that emerges with regard to the syllabic manipulation is a significant increase in sensitivity for complex syllables (i.e., CCV) relative to simple syllables (i.e., CV) in the TTGl, TTGm, TTSm, TTSl, STGm, SFp, SFa, PTa, and PTm. Furthermore, the SFa demonstrated a lateralization effect during the processing of syllabic information (the left SFa was sensitive to the syllabic manipulation but not the right SFa). Lastly, the PTm

Syllabic complexity

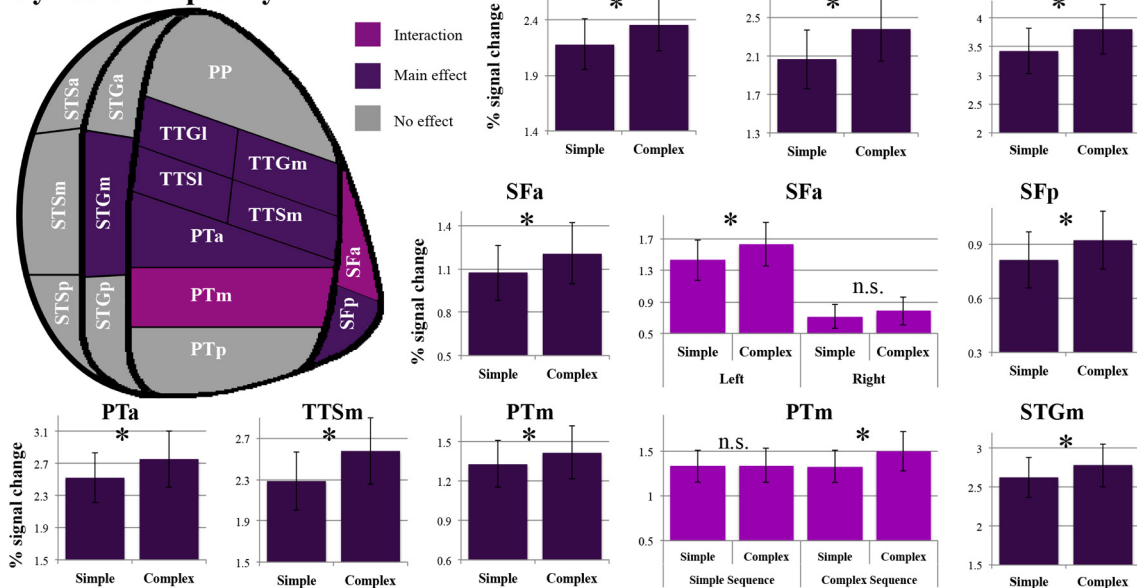


FIGURE 4 | Patterns of syllabic complexity effects observed in exploratory STP and STS ROI analysis. The results are shown on a flattened schematic representation of STP and STS showing the parcellation used in this study (different areas shown not to scale). Areas in dark purple exhibited a main effect of complexity and areas in lighter purple exhibited an interaction was observed (hemisphere * syllabic complexity for the SFa and syllabic complexity * supra-syllabic complexity for the PTm). Legend: PP, planum polare; TTG, transverse temporal gyrus (m, medial; l, lateral); TTS, transverse temporal sulcus (m, medial; l, lateral); PT, planum temporale (a, anterior; m, middle; p, posterior); SF, caudal Sylvian fissure (a, anterior; p, posterior); STG, superior temporal gyrus (a, anterior; m, middle; p, posterior); STS, superior temporal sulcus (a, anterior; m, middle; p, posterior); *significant contrast at $p_{FWE} = 0.05$, Bonferroni corrected; n.s. non-significant contrast. Error bars represent standard error of the mean.

planum polare; TTG, transverse temporal gyrus (m, medial; l, lateral); TTS, transverse temporal sulcus (m, medial; l, lateral); PT, planum temporale (a, anterior; m, middle; p, posterior); SF, caudal Sylvian fissure (a, anterior; p, posterior); STG, superior temporal gyrus (a, anterior; m, middle; p, posterior); STS, superior temporal sulcus (a, anterior; m, middle; p, posterior); *significant contrast at $p_{FWE} = 0.05$, Bonferroni corrected; n.s. non-significant contrast. Error bars represent standard error of the mean.

was the only region where an interaction between the syllabic and supra-syllabic manipulations was observed. In this region, the effect of syllabic complexity was restricted to complex sequences.

As shown in **Figure 5**, a main effect of supra-syllabic complexity was observed in the STSm [$F_{(1, 14)} = 5.89, p = 0.03$], the STGa [$F_{(1, 14)} = 5.39, p = 0.036$], the STGm [$F_{(1, 14)} = 27.38, p = 0.0001$], the PTa [$F_{(1, 14)} = 8.64, p = 0.01$], the TTSL [$F_{(1, 14)} = 10.95, p = 0.005$], the TTSM [$F_{(1, 14)} = 11.67, p = 0.004$], and the TTGm [$F_{(1, 14)} = 8.619, p = 0.011$]. We determined that for all seven regions, the complex sequences were driving the main effect of supra-syllabic complexity as they elicited higher levels of BOLD signal than simple sequences (paired sample t -tests, Bonferroni corrected). For the STSm and SFp, a hemisphere \times supra-syllabic interaction was observed [STSm: $F_{(1, 14)} = 10.06, p = 0.007$, SFp: $F_{(1, 14)} = 11.84, p = 0.004$]. For both regions, paired sample t -tests revealed that the source of the interaction was due to an effect of supra-syllabic complexity in the left hemisphere (STSm: $t = 3.851, p = 0.004$, SFp: $t = 2.55, p = 0.046$) but not the right hemisphere (STSm: $t = 0.64, p = 1$, SFp: $t = 0.965, p = 1$).

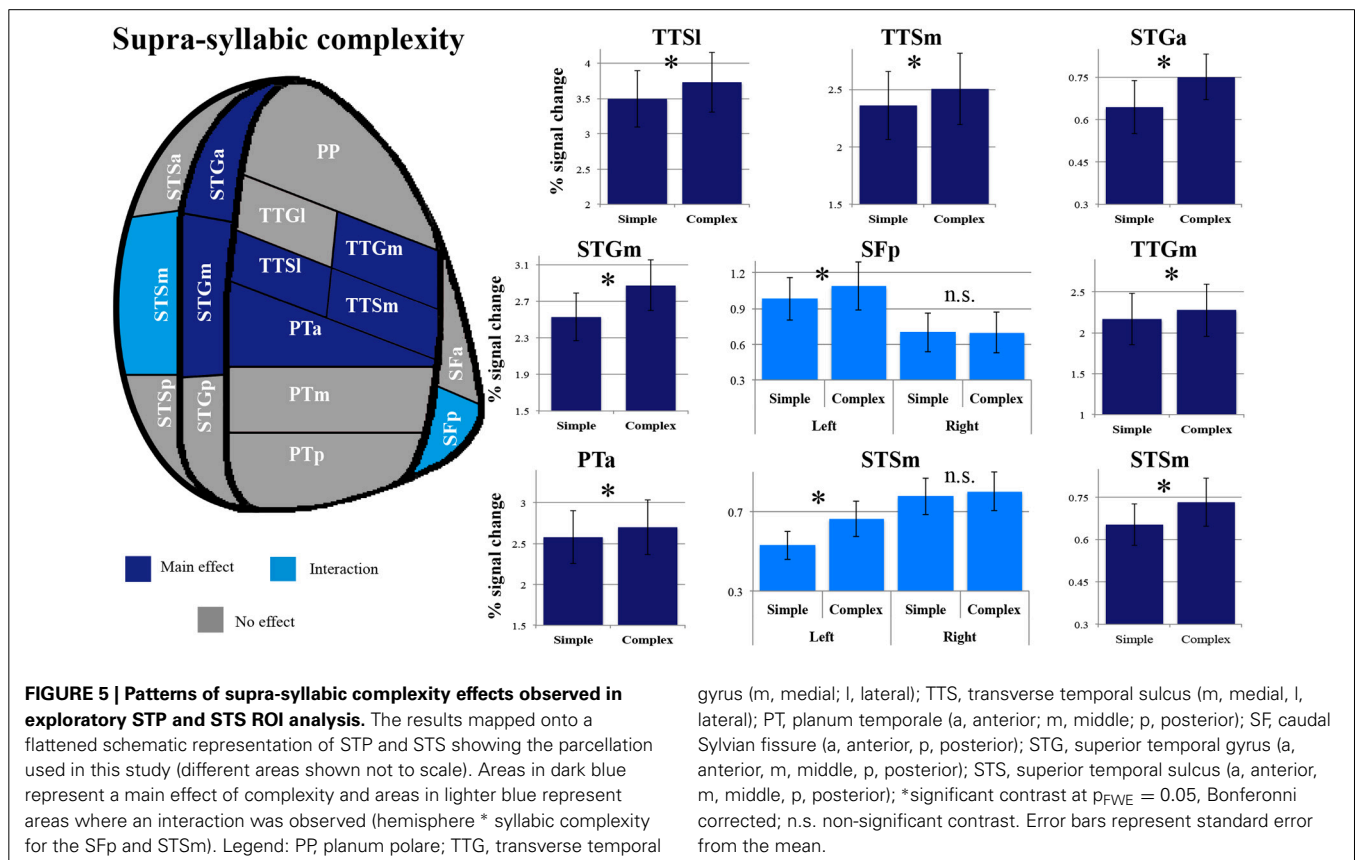
The overall pattern that emerges with regard to the supra-syllabic manipulation is a significant increase in sensitivity for complex sequences (i.e., three different syllables) relative to simple sequences (i.e., same syllable repeated $3 \times$) in the STSm, STGa, STGm, PTa, TTSL, TTSM, and TTGm. In addition, in two regions, the STSm and SFp an effect of hemisphere was observed. For both of these regions, the effect of supra-syllabic complexity was only observed in the left hemisphere.

In sum, the pattern that emerges from the ROI analysis suggest that some ROIs (STGm, TTSL, TTGm, TTSM, PTa) are sensitive to both experimental manipulations while others are only sensitive to one experimental manipulation (i.e., syllabic: left SFa, PTm, TTGL; supra-syllabic: left STSm, left SFp; for details refer to **Figure 6**). In addition, for ROIs that were sensitive to both manipulations, the magnitude of the manipulations was equivalent given the absence of syllabic complexity \times supra-syllabic complexity interaction within these regions.

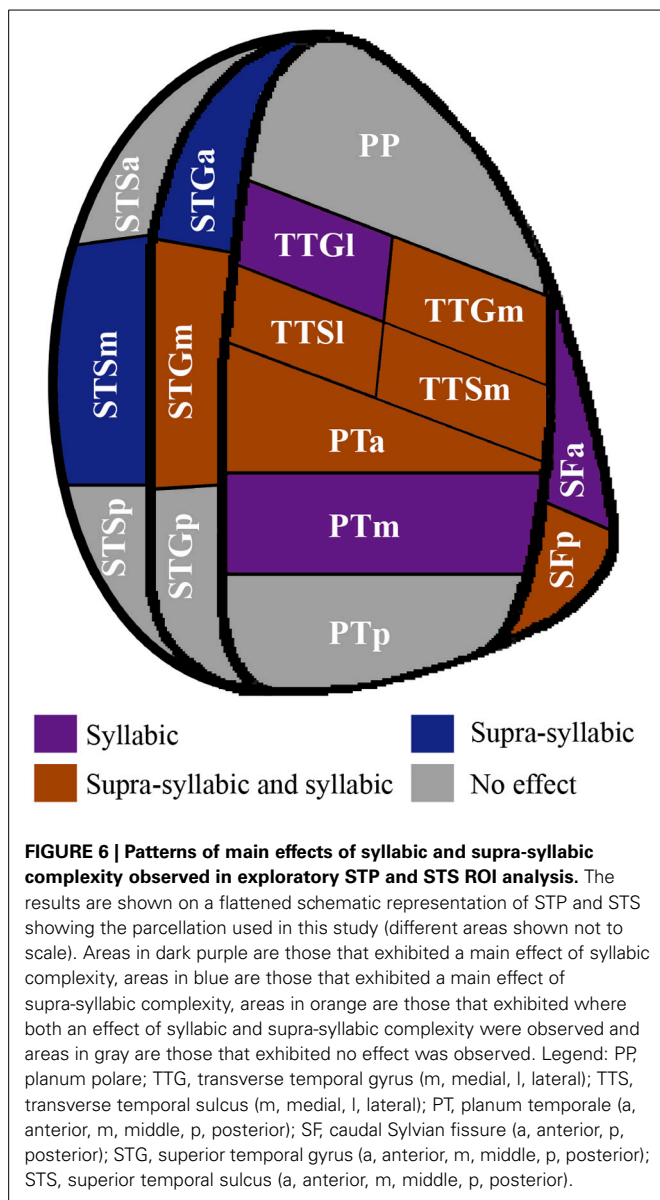
DISCUSSION

Neuroimaging studies have consistently documented the role of two large and functionally heterogeneous cortical areas, the STP and STS, in the perception of speech sounds. However, a detailed understanding of the role STP and STS in the processing of sub-lexical information has not yet emerged. This is largely related to the intrinsic complexity of the speech signal. Indeed, comprehending speech requires the interaction of complex sensory, perceptual, and cognitive mechanisms. The question, then, that naturally arises is whether these regions shows differential patterns of activation as of function of the type of information being processed (syllabic vs. supra-syllabic) (functional heterogeneity) and the specific sub-region (spatial heterogeneity).

The main objective of the current study was to examine, using fMRI, whether the processing of syllabic and supra-syllabic information during a passive listening task involve similar or distinct networks, with an emphasis on the STP and the STS. A passive listening paradigm was used in order to minimize task-related



gyrus (m, medial; l, lateral); TTS, transverse temporal sulcus (m, medial, l, lateral); PT, planum temporale (a, anterior; m, middle; p, posterior); SF, caudal Sylvian fissure (a, anterior, p, posterior); STG, superior temporal gyrus (a, anterior, m, middle, p, posterior); STS, superior temporal sulcus (a, anterior, m, middle, p, posterior); *significant contrast at $p_{FWE} = 0.05$, Bonferroni corrected; n.s. non-significant contrast. Error bars represent standard error from the mean.



cognitive/executive demands. Given the importance of the STP and STS in speech processing, we conducted an exploratory ROI analysis focusing on 16 bilateral STP/STS sub-regions to determine whether differential patterns of activation would be observed as a function of the type of information processed (i.e., syllabic or supra-syllabic). To preface the discussion, the results from the whole-brain analysis identified a network of regions involved in the perception of speech sounds that is consistent with previous neuroimaging studies that contrasted the processing of sublexical speech units to rest (Benson et al., 2001; Hugdahl et al., 2003; Wilson et al., 2004; Rimol et al., 2005; Wilson and Iacoboni, 2006). In addition, the results clearly demonstrate that the processing of auditory syllable sequences recruits both the left and right hemisphere, consistent with the notion that the processing of speech sounds is bilateral (Hickok and Poeppel, 2004, 2007; Hickok, 2009). The highly consistent results from the

whole-brain and ROI analysis demonstrate that both syllabic and supra-syllabic information are processed during passive listening. The anatomical specificity afforded by the ROI analyses allowed us to go further in exploring the specific functional contribution of sub-regions within the STP and STS during the perception of speech sounds. The findings are discussed below.

Results from the whole-brain analyses demonstrate widespread bilateral supratemporal activation resulting from the syllabic manipulation. The widespread extent of this activation was not expected based on previous fMRI results (McGettigan et al., 2011; Tremblay and Small, 2011). Of the few studies that have investigated the effect of consonant clusters during passive speech perception, in one study, activation within the right PT was scaled to syllabic complexity (Tremblay and Small, 2011) and in the other, no positive effect was reported (McGettigan et al., 2011). Our finding of widespread supratemporal effects may be related to the type of stimuli used. While in the present study we used meaningless sequences of syllables, Tremblay and Small (2011) used whole words, for which the mapping of sounds to linguistic representations may be more automatic, requiring less resources for the processing of syllabic information. However, if the processing of syllabic information interacts with lexical status, an effect of complexity should have been observed in the McGettigan et al. (2011) study given that pseudo-words were used, which are not overlearned stimuli with a stored lexical representation. It is possible that the absence of an effect of syllabic complexity in the latter study is attributable to a less salient experimental manipulation. In the present study, we contrasted sequences of syllables with either six or no consonant clusters, yielding a very robust effect. Although the use of a passive listening paradigm minimized attention-directed processes, mimicking more closely naturalistic speech perception situations, the use of syllable as experimental stimuli might have taxed to a greater extent phonological processes than the use of pseudo-words and words. This line of reasoning is consistent with neuropsychological and neurophysiological evidence suggesting that language comprehension does not depend on the processing of sublexical units (i.e., units smaller than words, such as syllables, phonemes, and phonetic features). For instance, it has been shown that patients with good word-level auditory comprehension abilities can fail on syllable and phoneme discrimination tasks (Basso et al., 1977; Boatman et al., 1995). Similarly, electrocortical mapping studies have provided evidence that phonological processes (e.g., syllable discrimination) and auditory word comprehension processes are not entirely circumscribed to the same STP regions (for a review, refer to: Boatman, 2004). In sum, while syllabic complexity effects are observed in sequences of syllables, further research need to determine whether and how syllabic information contributes to the perception of speech sounds and language comprehension.

Both whole-brain and exploratory ROI analyses identified a region that was sensitive to the presence or absence of consonant clusters; the lateral part of the primary auditory cortex (TTGI). In addition, the exploratory ROI analysis also identified the left SFa and PTm, as regions being sensitive to the syllabic manipulation. These results tentatively suggest that this effect stems from the addition of an extra consonant in the onset of the syllable and

not from differences between adjacent syllables (i.e., two different syllables). This pattern of response is consistent with the hypothesis that these regions are sensitive to the structure of the syllable (i.e., whether it is phonologically complex or not). Whether these regions respond to the complexity of the syllabic structure in general or to a specific component of the syllable (i.e., onset, rhyme, nucleus, or coda) however remains to be determined. Though the specific contribution of these three regions in the processing of syllabic information is still awaits further specifications, these three regions are nonetheless robustly activated during the perception of sublexical speech sounds (Benson et al., 2001; Hugdahl et al., 2003; Wilson et al., 2004; Rimol et al., 2005; Wilson and Iacoboni, 2006).

An alternative hypothesis that could explain the complexity effect related to the addition of a consonant to form a cluster is that these regions are responding to an increase in phonological working memory due to an increase in sequence length. This is because the addition of a consonant cluster to increase syllabic complexity also increases the length of the sequence. However, previous studies that have manipulated item length to investigate phonological working memory have reported mixed results that seem dependent upon (1) how length was modulated (CV-CCV vs. number of syllables), (2) the type of stimuli used (words, pseudowords), and (3) task demands (passive listening, judgment or naming task) (Okada et al., 2003; Strand et al., 2008; McGettigan et al., 2011). The most consistent finding is that stimulus length defined as the number of syllable yields more reliable results than the addition of consonant clusters. Moreover, if our syllabic manipulation results reflected an increase in phonological working memory, we would expect this contrast to yield clusters of activation within the pre-motor cortex, the IFG, and the IPL, that is, regions that are typically recruited during verbal working memory tasks (Paulesu et al., 1993; Honey et al., 2000; Marvel and Desmond, 2012). However, none of these regions was found in any of our contrasts.

Another alternative hypothesis is that the syllabic effect is due to an increase in acoustic/phonemic complexity. Indeed, consonant clusters are more complex than single consonants both acoustically and phonemically. Given that we parametrically varied both syllabic and supra-syllabic complexity, if this hypothesis were correct, we would expect the same regions to also exhibit an effect of supra-syllabic complexity since the presentation of three different syllables as opposed to the same syllable presented three times also increases acoustical complexity. In addition, we would also expect to see a syllabic complexity * supra-syllabic complexity interaction driven by a syllabic complexity effect for both simple and complex sequences and a stronger effect of syllabic complexity for the complex sequence. This pattern of result was not found in the SFa or the TTGl or the PTm. However, in the PTm, a region identical to the one reported by Tremblay and Small (2011), sensitivity to the syllabic manipulation was found only for the complex sequences. Combined with the observation that this region is involved in speech production (Dhanjal et al., 2008; Tourville et al., 2008; Peschke et al., 2009; Zheng et al., 2010) and that its activation magnitude varies as a function of syllabic complexity during both speech perception and production (Tremblay and Small, 2011), the result from the current study provides

additional support to the hypothesis that the right PT is involved in converting external auditory input into a phonological representation. Our results are in agreement with this hypothesis because an effect of syllabic complexity only emerged in this region when the sequences were composed of three different syllables (i.e., high supra-syllabic complexity). In itself, the addition of a consonant cluster increases the complexity of the syllable template. The additional complexity associated with processing three different sounds (high supra-syllabic complexity) enhances the syllabic manipulation, as three different consonant clusters have to be mapped onto phonological representations as opposed to three single consonants. In sum, the current results lend further support to the notion that regions within the posterior STP are important for the processing of phonological information, perhaps through a template matching mechanisms that uses spectrotemporal information to access stored syllabic representations (Griffiths and Warren, 2002; Warren et al., 2005).

Both the whole-brain and exploratory ROI analyses identified two regions, the STSm and STGa that were sensitive only to the supra-syllabic manipulation. This pattern of response suggests that these regions are involved in tracking changes that affect the structure of the sequence. In the present study, after having heard the second syllable of a sequence, participants could determine whether they would hear the same syllable again (i.e., in the case of simple sequences) or a different syllable (i.e., in the case of complex sequences). Thus, after the second syllable, for simple sequences the continuation was completely deterministic and prediction about upcoming sounds could be made. This pattern of response is also consistent with results from studies that have investigated the perception of speech sounds using a neural adaptation and oddball paradigm. In these studies, cluster of activation were observed within these regions in response to the presentation of a deviant stimulus (Vouloumanos et al., 2001; Joanisse et al., 2007). Overall, the results suggest that these regions are involved in representing sequences overtime. Thus, speech perception mechanisms, even in the absence of a task, are sensitive to changes that affect the structural properties of auditory sequences, consistent with previous work (Tremblay et al., 2012).

Both whole-brain and exploratory ROI analyses also identified a group of regions that was sensitive to both manipulations. These regions included the STGm, the TTGm, the TTSl, the TTSm, the PTa, and the SFp. Sensitivity to both manipulations suggests that these regions do not exhibit a differentiation in processing syllabic or supra-syllabic information. In a previous neuroimaging study using the same parcellation scheme of the STP, both the TTSl and PTa responded to speech and non-speech sounds, whereas the STGm, SFp, and TTGm exhibited an absolute preference for speech sounds (Tremblay et al., 2012), consistent with the idea that regions located anterior and lateral the primary auditory cortex are involved in processing changes in spectro-temporal features (Scott and Johnsrude, 2003). These results suggest that both syllabic and supra-syllabic information recruits common mechanisms involved in processing acoustical information.

In the current study, we explored the neural mechanisms involved in the processing of syllabic and supra-syllabic information during passive speech perception. We demonstrated that both syllabic and supra-syllabic information are processed

automatically during passive speech listening, a finding that is consistent with the finding of distinct neural representations for syllable and sequence-level information during speech production (Bohland and Guenther, 2006; Peeva et al., 2010). Importantly, these findings suggest that processing of sublexical information is automatic, at least during the processing of meaningless syllable sequences. Future studies need to examine whether the processing of sub-lexical information is automatic and necessary during language comprehension using more naturalistic stimuli such as words or connected speech. It is possible that the recruitment of phonological mechanisms depends upon the context, or the kind or quality of auditory stimuli being processed. Degraded speech stimuli, for instance, could recruit sublexical phonological mechanisms to a greater extent than high-quality speech sounds. Nevertheless, the present study offers new insight into the functional neuroanatomy of the system involved in sublexical phonological processing, highlighting the importance of the anterior two-thirds of the PT, the primary auditory cortices and the middle part of the STS and STG in these processes.

ACKNOWLEDGMENTS

This work was supported by research grants from the *Natural Sciences and Engineering Research Council of Canada (NSERC, grant #1958126)* and from the *Fonds de la Recherche en Santé Mentale de Québec—Société et Culture (FRQSC, grant #169309)* to Pascale Tremblay. Support from the *Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec* is also gratefully acknowledged. Technical support was provided by the “*Consortium d'imagerie en neurosciences et santé mentale de Québec*” (CINQ) for protocol development and MRI data acquisition.

REFERENCES

- Argall, B. D., Saad, Z. S., and Beauchamp, M. S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. *Hum. Brain Mapp.* 27, 14–27. doi: 10.1002/hbm.20158
- Basso, A., Casati, G., and Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex* 13, 85–95. doi: 10.1016/S0010-9452(77)80057-9
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. Ser. B* 57, 289–300.
- Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang.* 78, 364–396. doi: 10.1006/brln.2001.2484
- Bilodeau-Mercure, M., Lortie, C. L., Sato, M., Guitton, M. J., and Tremblay, P. (2014). The neurobiology of speech perception decline in aging. *Brain Struct. Funct.* doi: 10.1007/s00429-013-0695-3. [Epub ahead of print].
- Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., and Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain* 119(pt 4), 1239–1247.
- Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., et al. (1994). Functional magnetic resonance imaging of human auditory cortex. *Ann. Neurol.* 35, 662–672. doi: 10.1002/ana.410350606
- Boatman, D. (2004). Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* 92, 47–65. doi: 10.1016/j.cognition.2003.09.010
- Boatman, D., Lesser, R. P., and Gordon, B. (1995). Auditory speech processing in the left temporal lobe: an electrical interference study. *Brain Lang.* 51, 269–290. doi: 10.1006/brln.1995.1061
- Bohland, J. W., and Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage* 32, 821–841. doi: 10.1016/j.neuroimage.2006.04.173
- Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2002). Functional anatomy of intra- and cross-modal lexical tasks. *Neuroimage* 16, 7–22. doi: 10.1006/nimg.2002.1081
- Buchman, A. S., Garron, D. C., Trost-Cardamone, J. E., Wichter, M. D., and Schwartz, M. (1986). Word deafness: one hundred years later. *J. Neurol. Neurosurg. Psychiatry* 49, 489–499. doi: 10.1136/jnnp.49.5.489
- Burton, M. W., Locasto, P. C., Krebs-Noble, D., and Gullapalli, R. P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage* 26, 647–661. doi: 10.1016/j.neuroimage.2005.02.024
- Burton, M. W., Small, S. L., and Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12, 679–690. doi: 10.1162/089892900562309
- Celsis, P., Boulouvar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., et al. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9, 135–144. doi: 10.1006/nimg.1998.0389
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014
- Dahl, A., Omdal, R., Waterloo, K., Joakimsen, O., Jacobsen, E. A., Koldingsnes, W., et al. (2006). Detection of cerebral embolic signals in patients with systemic lupus erythematosus. *J. Neurol. Neurosurg. Psychiatry* 77, 774–779. doi: 10.1136/jnnp.2005.084285
- Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194. doi: 10.1006/nimg.1998.0395
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33. doi: 10.1016/j.neuroimage.2004.09.039
- Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain* 115(pt 6), 1753–1768.
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Dhanjal, N. S., Handunnetthi, L., Patel, M. C., and Wise, R. J. (2008). Perceptual systems controlling speech production. *J. Neurosci.* 28, 9969–9975. doi: 10.1523/JNEUROSCI.2607-08.2008
- Dhankhar, A., Wexler, B. E., Fulbright, R. K., Halwes, T., Blamire, A. M., and Shulman, R. G. (1997). Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rates. *J. Neurophysiol.* 77, 476–483.
- Duvernoy, H. M. (1991). *The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI, and Blood Supply*. New York, NY: Springer-Wien.
- Eden, G. E., Joseph, J. E., Brown, H. E., Brown, C. P., and Zeffiro, T. A. (1999). Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. *Magn. Reson. Med.* 41, 13–20.
- Edmister, W. B., Talavage, T. M., Ledden, P. J., and Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Hum. Brain Mapp.* 7, 89–97.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/S0896-6273(02)00569-X
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. doi: 10.1006/nimg.1998.0396
- Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., et al. (2004). Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22. doi: 10.1093/cercor/bhg087
- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037

- Giraud, A. L., and Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *J. Cogn. Neurosci.* 13, 754–765. doi: 10.1162/08989290152541421
- Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., and Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. *Brain Res.* 1515, 55–65. doi: 10.1016/j.brainres.2013.03.024
- Gracco, V. L., Tremblay, P., and Pike, B. (2005). Imaging speech production using fMRI. *Neuroimage* 26, 294–301. doi: 10.1016/j.neuroimage.2005.01.033
- Griffiths, T. D., and Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353. doi: 10.1016/S0166-2236(02)02191-4
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliot, M. R., et al. (1999). Sparse temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7, 213–223.
- Hickok, G. (2009). The functional neuroanatomy of language. *Phys. Life Rev.* 6, 121–143. doi: 10.1016/j.plrev.2009.06.001
- Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Honey, G. D., Bullmore, E. T., and Sharma, T. (2000). Prolonged reaction time to a verbal working memory task predicts increased power of posterior parietal cortical activation. *Neuroimage* 12, 495–503. doi: 10.1006/nimg.2000.0624
- Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L. M., and Niemi, J. (2003). The effects of attention on speech perception: an fMRI study. *Brain Lang.* 85, 37–48. doi: 10.1016/S0093-934X(02)00500-X
- Jacquemot, C., Pallier, C., Lebihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 23, 9541–9546.
- Joanisse, M. F., and Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage* 19, 64–79. doi: 10.1016/S1053-8119(03)00046-6
- Joanisse, M. F., Zevin, J. D., and McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cereb. Cortex* 17, 2084–2093. doi: 10.1093/cercor/bhl124
- Johnstone, T., Ores Walsh, K. S., Greischar, L. L., Alexander, A. L., Fox, A. S., Davidson, R. J., et al. (2006). Motion correction and the use of motion covariates in multiple-subject fMRI analysis. *Hum. Brain Mapp.* 27, 779–788. doi: 10.1002/hbm.20219
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., and Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621–1631. doi: 10.1093/cercor/bhi040
- Liebenthal, E., Sabri, M., Beardsley, S. A., Mangalathu-Arumana, J., and Desai, A. (2013). Neural dynamics of phonological processing in the dorsal auditory stream. *J. Neurosci.* 33, 15414–15424. doi: 10.1523/JNEUROSCI.1511-13.2013
- Marvel, C. L., and Desmond, J. E. (2012). From storage to manipulation: how the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain Lang.* 120, 42–51. doi: 10.1016/j.bandl.2011.08.005
- McGettigan, C., Warren, J. E., Eisner, F., Marshall, C. R., Shanmugalingam, P., and Scott, S. K. (2011). Neural correlates of sublexical processing in phonological working memory. *J. Cogn. Neurosci.* 23, 961–977. doi: 10.1162/jocn.2010.21491
- Meltzer, J. A., Negishi, M., and Constable, R. T. (2008). Biphasic hemodynamic responses influence deactivation and may mask activation in block-design fMRI paradigms. *Hum. Brain Mapp.* 29, 385–399. doi: 10.1002/hbm.20391
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x
- Newman, S. D., and Twieg, D. (2001). Differences in auditory processing of words and pseudowords: an fMRI study. *Hum. Brain Mapp.* 14, 39–47. doi: 10.1002/hbm.1040
- Obleser, J., Zimmermann, J., Van Meter, J., and Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb. Cortex* 17, 2251–2257. doi: 10.1093/cercor/bhl133
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex* 20, 2486–2495. doi: 10.1093/cercor/bhp318
- Okada, K., Smith, K. R., Humphries, C., and Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport* 14, 2323–2326. doi: 10.1097/00001756-200312190-00007
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Paulesu, E., Frith, C. D., and Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345. doi: 10.1038/362342a0
- Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J. L., Nazarian, B., et al. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage* 50, 626–638. doi: 10.1016/j.neuroimage.2009.12.065
- Peschke, C., Ziegler, W., Kappes, J., and Baumgaertner, A. (2009). Auditory-motor integration during fast repetition: the neuronal correlates of shadowing. *Neuroimage* 47, 392–402. doi: 10.1016/j.neuroimage.2009.03.061
- Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain Lang.* 55, 317–351. discussion 352–385.
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso Del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7865–7870. doi: 10.1073/pnas.0509989103
- Rimol, L. M., Specht, K., Weis, S., Savoy, R., and Hugdahl, K. (2005). Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. *Neuroimage* 26, 1059–1067. doi: 10.1016/j.neuroimage.2005.03.028
- Saad, Z. S., Glen, D. R., Chen, G., Beauchamp, M. S., Desai, R., and Cox, R. W. (2009). A new method for improving functional-to-structural MRI alignment using local Pearson correlation. *Neuroimage* 44, 839–848. doi: 10.1016/j.neuroimage.2008.09.037
- Saad, Z. S., Reynolds, R. C., Argall, B. D., Japee, S., and Cox, R. W. (2004). “SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI,” in *Proceedings of the IEEE International Symposium on Biomedical Imaging* (Arlington, TX), 1510–1513.
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400
- Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107. doi: 10.1016/S0166-2236(02)00037-1
- Stach, B. A. (2010). *Clinical Audiology: An Introduction*. Clifton Park, NY: Delmar Cengage Learning.
- Strand, F., Forssberg, H., Klingberg, T., and Norrelgen, F. (2008). Phonological working memory with auditory presentation of pseudo-words – an event related fMRI study. *Brain Res.* 1212, 48–54. doi: 10.1016/j.brainres.2008.02.097
- Tanaka, Y., Yamadori, A., and Mori, E. (1987). Pure word deafness following bilateral lesions. A psychophysical analysis. *Brain* 110(pt 2), 381–403.
- Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054
- Tremblay, P., Baroni, M., and Hasson, U. (2012). Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. *Neuroimage* 66C, 318–332. doi: 10.1016/j.neuroimage.2012.10.055
- Tremblay, P., Deschamps, I., and Gracco, V. L. (2013). Regional heterogeneity in the processing and the production of speech in the human planum temporale. *Cortex* 49, 143–157. doi: 10.1016/j.cortex.2011.09.004
- Tremblay, P., and Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *Neuroimage* 57, 1561–1571. doi: 10.1016/j.neuroimage.2011.05.067
- Vouloumanos, A., Kiehl, K. A., Werker, J. F., and Liddle, P. F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.* 13, 994–1005. doi: 10.1162/089892901753165890
- Warren, J. E., Wise, R. J., and Warren, J. D. (2005). Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci.* 28, 636–643. doi: 10.1016/j.tins.2005.09.010
- Wilson, S. M., and Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* 33, 316–325. doi: 10.1016/j.neuroimage.2006.05.032

- Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702. doi: 10.1038/nn1263
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., et al. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatr. Res.* 17, 37–49. doi: 10.1016/0022-3956(82)90033-4
- Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846–849. doi: 10.1126/science.1589767
- Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., and Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS ONE* 6:e20963. doi: 10.1371/journal.pone.0020963
- Zheng, Z. Z., Munhall, K. G., and Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *J. Cogn. Neurosci.* 22, 1770–1781. doi: 10.1162/jocn.2009.21324

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 February 2014; accepted: 17 June 2014; published online: 08 July 2014.

Citation: Deschamps I and Tremblay P (2014) Sequencing at the syllabic and supra-syllabic levels during speech perception: an fMRI study. *Front. Hum. Neurosci.* 8:492. doi: 10.3389/fnhum.2014.00492

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Deschamps and Tremblay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences

Sylvia Vitello^{1*}, Jane E. Warren^{1,2}, Joseph T. Devlin¹ and Jennifer M. Rodd¹

¹ Department of Experimental Psychology, University College London, London, UK

² Department of Language and Communication, University College London, London, UK

Edited by:

Patti Adank, University College London, UK

Reviewed by:

Stefanie E. Kuchinsky, University of Maryland, USA

Glyn Paul Hallam, The University of York, UK

*Correspondence:

Sylvia Vitello, Department of Experimental Psychology, University College London, Gower Street, London, WC1E 6BT, UK
e-mail: s.vitello@ucl.ac.uk

Semantic ambiguity resolution is an essential and frequent part of speech comprehension because many words map onto multiple meanings (e.g., “bark,” “bank”). Neuroimaging research highlights the importance of the left inferior frontal gyrus (LIFG) and the left posterior temporal cortex in this process but the roles they serve in ambiguity resolution are uncertain. One possibility is that both regions are engaged in the processes of semantic reinterpretation that follows incorrect interpretation of an ambiguous word. Here we used fMRI to investigate this hypothesis. 20 native British English monolinguals were scanned whilst listening to sentences that contained an ambiguous word. To induce semantic reinterpretation, the disambiguating information was presented after the ambiguous word and delayed until the end of the sentence (e.g., “the teacher explained that the BARK was going to be very damp”). These sentences were compared to well-matched unambiguous sentences. Supporting the reinterpretation hypothesis, these ambiguous sentences produced more activation in both the LIFG and the left posterior inferior temporal cortex. Importantly, all but one subject showed ambiguity-related peaks within both regions, demonstrating that the group-level results were driven by high inter-subject consistency. Further support came from the finding that activation in both regions was modulated by meaning dominance. Specifically, sentences containing biased ambiguous words, which have one more dominant meaning, produced greater activation than those with balanced ambiguous words, which have two equally frequent meanings. Because the context always supported the less frequent meaning, the biased words require reinterpretation more often than balanced words. This is the first evidence of dominance effects in the spoken modality and provides strong support that frontal and temporal regions support the updating of semantic representations during speech comprehension.

Keywords: fMRI, LIFG, lexical ambiguity, speech comprehension, semantics, reinterpretation, sentence processing

INTRODUCTION

Many of the words encountered in everyday language have multiple meanings, which makes the process of mapping word form onto meaning often ambiguous. This means that listeners must routinely combine various kinds of contextual information to understand the meaning that is intended by the speaker. For example, to understand the sentence “the woman used a microphone to make the toast,” listeners must use the word “microphone” to understand that the semantically ambiguous word “toast” refers to a celebratory speech rather than grilled bread. Importantly, such ambiguity is often not noticed by listeners (Rodd et al., 2005), suggesting that disambiguation is generally a highly efficient and effective process. An understanding of the neural substrates supporting this process is essential in order to gain insight into the efficiency of language comprehension and because the breakdown of this process can lead to severe communication difficulties due to the prominence of ambiguous words in everyday language (Parks et al., 1998; Rodd et al., 2002).

Cumulative evidence from recent neuroimaging studies has highlighted the importance of two brain areas for semantic ambiguity resolution: the left inferior frontal gyrus (IFG) and the left posterior temporal cortex (Rodd et al., 2005, 2010b, 2012; Davis et al., 2007; Mason and Just, 2007; Zempleni et al., 2007; Bekinschtein et al., 2011). However, the relative contributions of these regions to ambiguity processing are uncertain. Psycholinguistic research converges on several cognitive processes that underpin semantic ambiguity resolution: accessing the alternative meanings of an ambiguous word, selecting a single meaning, and reinterpreting that meaning when an incorrect selection is initially made (e.g., Duffy et al., 1988; Gernsbacher, 1991; Simpson, 1994; Twilley and Dixon, 2000; Duffy et al., 2001; Rodd et al., 2010a).

One hypothesis of the contribution of LIFG and posterior temporal cortex to ambiguity resolution is that both regions play an important role in reinterpretation processes (Zempleni et al., 2007; Rodd et al., 2010a,b; Bekinschtein et al., 2011). Semantic reinterpretation occurs when listeners encounter context that is

not consistent with their initial understanding of the ambiguous word, requiring them to suppress the initially-selected meaning and integrate the alternative, contextually-appropriate interpretation (e.g., Duffy et al., 1988; Twilley and Dixon, 2000; Rodd et al., 2010a). Evidence for the reinterpretation hypothesis comes from the finding that activation in these frontal and temporal regions are greater for sentences with a higher likelihood of reinterpretation. For example, several functional MRI (fMRI) studies have shown increased activation in these regions for sentences in which the disambiguating information is delayed until after the ambiguous word (e.g., “The teacher explained that the BARK was going to be very damp”) compared to unambiguous sentences (Mason and Just, 2007; Zempleni et al., 2007; Rodd et al., 2010b, 2012; Bekinschtein et al., 2011). These are known as late-disambiguation sentences. Delaying the disambiguating information makes it impossible for listeners to determine the intended (e.g., tree) meaning of the ambiguous word when it is initially encountered. Thus, listeners will initially misinterpret the correct meaning on some occasions (i.e., first selecting the dog meaning of “bark”) and need to revise their understanding when they encounter the disambiguating information later on in the sentence (i.e., to the tree meaning). This process of initial meaning selection followed by reinterpretation is assumed in many influential cognitive models of semantic ambiguity resolution (Swinney, 1979; Twilley and Dixon, 2000; Duffy et al., 2001), on the basis of numerous cross-modal priming studies and eye-movement research which show that listeners and readers select a meaning within a few hundred millisecond of encountering an ambiguous word (e.g., Swinney, 1979; Seidenberg et al., 1982; Rayner and Duffy, 1986; Duffy et al., 1988). Various psycholinguistic studies, including eye-movement and dual-task research, also provide converging evidence that reinterpretation occurs for late-disambiguation sentences by showing that listeners and readers incur greater behavioral costs of processing the disambiguating regions in these sentences (e.g., longer reading times or poorer performance on an unrelated concurrent task) compared to processing equivalent regions in early-disambiguation sentences (e.g., “The hunter thought that the HARE in the field was actually a rabbit”) or unambiguous sentences (Rayner and Duffy, 1986; Duffy et al., 1988, 2001; Rodd et al., 2010a). Zempleni et al. (2007) provide more direct support that frontal and temporal regions support such reinterpretation processes by showing that activation in the LIFG and posterior middle/inferior temporal gyrus was modulated by meaning dominance, that is, how frequent the intended meaning is relative to the other meanings. Specifically activation in these regions was greater for late-disambiguation sentences that corresponded to the subordinate (i.e., less frequent) meaning than the dominant meaning (Zempleni et al., 2007). Reinterpretation is more likely in subordinate-biased sentences because people will typically select the dominant meaning in the absence of prior biasing context (Rayner and Duffy, 1986; Duffy et al., 1988; Simpson and Krueger, 1991).

Rodd et al. (2012) further proposed that the LIFG, in particular, may also be important for the initial selection of an ambiguous word’s meaning that occurs when the word is initially encountered during a sentence (Twilley and Dixon, 2000). This suggestion was based on the finding that the LIFG, but not the

posterior temporal cortex, was also more active for sentences in which reinterpretation was unlikely (compared to unambiguous sentences). These were sentences in which the disambiguating information *preceded* the ambiguous word. In addition, this region also showed activation that was temporally associated with the ambiguous word as well as the disambiguating information in late-disambiguation sentences. Thus, these results suggested that the LIFG may be involved in multiple processes of ambiguity resolution, not only when a meaning needs to be reinterpreted.

Supporting evidence for the involvement of the LIFG and posterior temporal cortex in reinterpretation and/or initial meaning selection, however, is not conclusive on various levels. First, the functional contributions of these regions to these two ambiguity-related processes is uncertain because not all studies have found the same response pattern to the different types of ambiguous sentences that load on these processes (Mason and Just, 2007; Bekinschtein et al., 2011). Second, different methods for examining neural responses to reinterpretation and/or initial meaning selection have been used for written sentences compared to spoken sentences. For example, studies have assessed how meaning dominance modulates ambiguity-related neural responses but these have only been conducted on visually-presented late-disambiguation sentences (Mason and Just, 2007; Zempleni et al., 2007). It is important to examine whether such dominance patterns found in the visual modality also replicate for spoken sentences in order to understand whether these ambiguity-related responses generalize across modalities. Although many of the regions reported in semantic ambiguity studies are considered modality-general (Binder et al., 2009; Novick et al., 2009; Price, 2012), it is possible that speech places different demands on ambiguity processes (particularly working memory aspects) due to the transient, fast-fading nature of the speech signal and, thus, may place different demands on the underlying neural circuitry. Third, the precise nature of the LIFG and posterior temporal cortex’s involvement in ambiguity processes is also uncertain because there is considerable anatomical variability in the locus and extent of the ambiguity responses in these regions reported across studies (Rodd et al., 2005, 2010b, 2012; Mason and Just, 2007; Zempleni et al., 2007; Bekinschtein et al., 2011). As these anatomical differences relate to different anatomical regions that have been associated with different functions (see Price, 2012, for a recent review), it is important to explore the potential sources of this variability. It is possible that such variability reflects effects of statistical thresholds, differences in ambiguous stimuli or experimental protocols or even inter-subject functional variability given the finding of looser function-anatomy mappings for high-level cognitive processes (Duncan et al., 2009; Tahmasebi et al., 2012).

Furthermore, it is unclear how these ambiguity-responsive regions relate to those associated with sentence comprehension more generally. Do semantically ambiguous words place additional demands on regions that are already involved in the processing of sentences in general or do they engage regions that are more specific to semantically demanding stimuli? Neural models of language comprehension give different answers to this question. For example, Hagoort’s unification account of LIFG function argues for the former, imputing a sentence-general function

to this region (Hagoort, 2005, 2013), while Novick and colleagues’ conflict resolution account argues for the latter (Novick et al., 2005, 2009). Such differences in perspective are also found across theories of the posterior temporal cortex’s function in language processing (e.g., Hickok and Poeppel, 2007; Jefferies, 2013).

In summary, the current literature raises several questions regarding the involvement of the LIFG and posterior temporal cortex in semantic ambiguity resolution. What functional roles do these regions play in ambiguity resolution and sentence comprehension more generally? Which specific anatomical sub-fields within these regions are engaged by semantic ambiguity? How consistent is this ambiguity network across individuals? These questions were investigated using fMRI. Neural responses to a large set of late-disambiguation sentences were compared with those to well-matched unambiguous sentences. Based on previous research, it was predicted that ambiguity-elevated responses would be broadly found in the LIFG and the left posterior temporal cortex (Rodd et al., 2005, 2010b, 2012; Davis et al., 2007; Mason and Just, 2007; Zempleni et al., 2007; Bekinschtein et al., 2011). The areas showing a significant ambiguity effect were then investigated to answer three specific questions pertaining to function and inter-subject variability:

1. Are activations within these regions specific to ambiguous sentences or present for all sentences, albeit to a greater extent during ambiguous sentences? To address this question, unambiguous sentences were compared to a low level baseline condition. If regions showing an ambiguity effect support operations that are routinely involved in sentence comprehension then they should also show greater activation for unambiguous sentences compared to baseline. If, however, they support processes that are more specific to ambiguous stimuli, then they should not show an increased response to unambiguous sentences.
2. Is it possible to separate semantic reinterpretation processes from initial meaning selection in these regions by considering meaning dominance and are dominance effects for spoken sentences similar to those found for visually-presented sentences? Here dominance refers to the fact that some ambiguous words are biased in terms of the relative frequencies of their alternative meanings whereas others are balanced. Biased words such as “bank” have one more dominant meaning (i.e. financial institution rather than side of a river) whereas balanced words such as “bark” have meanings that are relatively equal in frequency (i.e., dog vs. tree). In this experiment, biased sentences were always disambiguated toward their subordinate meaning. By comparing responses to biased and balanced ambiguous words, we investigated the two key functions of ambiguity resolution: semantic reinterpretation and initial meaning selection. If regions showing an ambiguity effect are primarily involved in semantic reinterpretation, then responses should be greater for biased than balanced sentences because they have a stronger frequency bias that makes the inappropriate (dominant) meaning likely to be initially selected on most trials and, thus, need reinterpretation. In contrast, any regions that are primarily involved in initial meaning selection should show greater activation for balanced

than biased sentences because this process is more difficult since listeners have less strong preferences for one particular meaning of these words (Duffy et al., 1988, 2001; Twilley and Dixon, 2000). If regions are involved in both processes (relatively equally), then they may show equivalently strong activation to both balanced and biased sentences, since they both load on (at least) one of these processes. In addition, two types of biased sentences were compared: strongly subordinate and weakly subordinate words. This comparison allows us to examine whether responses are merely related to the likelihood of reinterpretation, where the dominance pattern would be: “strongly biased” > “weakly biased” > “balanced,” or whether a less linear relationship exists between reinterpretation and ambiguity-responses. For example, a region may be especially engaged when very infrequent meanings need to be integrated, which would produce a pattern of: “strongly biased” > “weakly biased” = “balanced”. See Table 1 for example sentences in each of the ambiguous and unambiguous conditions.

3. Finally, how consistent are these neuronal ambiguity-effects across individuals? Inter-subject variability of regions showing an ambiguity effect were assessed by examining whether the regions that showed reliable activation at the group-level were activated in all subjects.

METHODS
PARTICIPANTS

Twenty native monolingual British English speakers (11 female), aged 18–35 ($M = 23.8$) participated in the study. All were right-handed, had normal or corrected-to-normal vision and had no known hearing or language impairment. Participants were recruited from the UCL experimental subject pool and were paid for their participation. All gave informed consent and appropriate ethical approval was obtained from the UCL Departmental Ethics committee.

STIMULI

Ninety two ambiguous auditory sentences were created based on items from Rodd et al. (2010a). They were all late-disambiguation sentences where the disambiguating context was presented after a semantically ambiguous noun. For example, in the sentence “the woman had to make the *toast* with a very old *microphone*,” “toast” is the ambiguous word (i.e., grilled bread vs. celebratory speech)

Table 1 | Sentence conditions.

Sentence condition	
Ambiguous (strongly-biased)	(e.g., The woman had to make the TOAST with a very old <u>microphone</u>)
Ambiguous (weakly-biased)	(e.g., The man was told that an ORGAN was not available for the <u>choir</u>)
Ambiguous (balanced)	(e.g., The teacher explained that the BARK was going to be very <u>damp</u>)
Unambiguous (control)	(e.g., The teacher explained that the steam was going to be very hot)

In each example, the ambiguous word is capitalized and the disambiguating word is underlined.

and “microphone” is the disambiguating word. On average, the ambiguous words were presented 6.70 ($SD = 1.00$) words into the sentence and were a mixture of homographic and heterographic nouns (e.g., bark, night/knight). The disambiguation was always provided by the sentence-final word, except in four sentences where it was the last two words. There were at least 4 words between the ambiguous and disambiguating words ($M = 5.79$, $SD = 1.46$) to give listeners enough time to select their preferred meaning before they hear the disambiguating information (e.g., Swinney, 1979; Seidenberg et al., 1984; Rodd et al., 2010a). To ensure that the rest of the words were neutral, the sentences were created such that only the disambiguating word needed to be changed to instigate the alternative meaning. For example, in the above example, “microphone” could be replaced by “grill.” These alternative versions were not employed in the experiment.

To elicit semantic reinterpretation, the disambiguating words were chosen to correspond to the less frequent meaning of the ambiguous word because psycholinguistic research demonstrates that the dominant, (and in this case, incorrect), meaning is usually initially selected prior to disambiguating information (Duffy et al., 1988; Simpson and Krueger, 1991; Twilley and Dixon, 2000). The subordinate meaning was based on pre-test scores obtained by Rodd et al. (2010a). To validate these preferences, an independent group of 59 participants performed an extended version of the standard word association task typically used to measure meaning preferences (Twilley et al., 1994). Each ambiguous word was presented in isolation (e.g., fan) and participants typed the first related word that came to mind (e.g., wind, follower, cool). Because some responses could relate to more than one meaning (e.g., cool), after all the isolated words had been presented, the participants selected a definition of their intended meaning (e.g., admirer vs. ventilation device). This ensured that equivocal responses could be coded accurately (for further details see Vitello, 2014). A dominance score was subsequently calculated as the proportion of codable responses that were consistent with the meaning used in the experimental sentence (minimum 31 data-points per item). As expected, most words had low dominance scores ($M = 0.25$, $SD = 0.20$) indicating that the meaning used in the experimental sentences was the less preferred, infrequent meaning for the majority of items. These scores spanned across the four main categories of meaning dominance reported in the psycholinguistic ambiguity literature (Rayner and Duffy, 1986; Duffy et al., 1988; Sereno, 1995; Vuong and Martin, 2011): (1) 32 words were strongly subordinate-biased, where the meaning used in the experimental sentences was very infrequent, on average, preferred by only 6% of listeners (dominance range: 0–0.14); (2) 27 words were weakly subordinate-biased, where the sentence meaning was fairly infrequent, on average preferred by 21% of listeners (dominance range: 0.16–0.30); (3) 27 words were balanced, where the sentence meaning was one of two (or more) relatively equally frequent meanings, on average, preferred by 39% of listeners (dominance range: 0.31–0.54); (4) the remaining six sentences had high dominance scores, where the sentence meaning was, on average preferred by 77% of listeners (dominance range = 0.65–0.84). The range of “balanced” scores coheres with studies in which the less likely meaning of the balanced words was chosen (Rayner and Duffy, 1986; Sereno, 1995; Vuong

and Martin, 2011). One-way independent-measures ANOVAs showed that the strongly-biased, weakly-biased and balanced conditions were matched on sentence-level properties (duration in seconds, number of syllables, number of words, position of the ambiguous word, position of the disambiguating word, naturalness rating; all $ps > 0.2$) as well as on lexical properties of the ambiguous word [log-transformed frequency, number of letters, number of meanings and number of senses, where “meanings” refers to semantically and etymologically unrelated meanings (e.g., bark) and “word senses” are semantically related (e.g., run), Rodd et al., 2002, all $ps > 0.09$].

Each ambiguous sentence was paired with a well-matched unambiguous sentence of similar syntactic structure that had a low-ambiguity noun in the position of the ambiguous word. For example, “the student had to wrap the *wrist* with a very old bandage.” Statistical tests confirmed that the ambiguous words had significantly more meanings and senses than the unambiguous words [$t_{(91)} = 8.14$, $p < 0.001$; $t_{(91)} = 8.31$, $p < 0.001$, respectively] (Online Wordsmyth English Dictionary-Thesaurus, Parks et al., 1998). They, however, did not differ significantly on other lexical properties, including log-transformed word frequency (CELEX lexical database, Baayen et al., 1995), and number of letters (all $ps > 0.1$). See **Table 2** for corresponding descriptive statistics. The sets of ambiguous and unambiguous sentences were additionally matched on physical duration, number of syllables and number of words (all $ps > 0.1$). On average, both sets were also judged as highly natural, although statistically, the ambiguous sentences had lower naturalness ratings when rated on a 1 (highly unnatural) to 7 (highly natural) point scale by an independent group of 15 participants [$t_{(91)} = 3.98$, $p < 0.001$]. See **Table 2** for sentence-level descriptive statistics. All sentences were spoken by the same female speaker (JMR).

Additionally, 46 filler sentences (50% ambiguous) were employed with the same structure as the experimental sentences. 14 were used in an initial practice block, 24 were catch sentences and the remaining 8 constituted dummy trials at the beginning of the fMRI runs. Catch sentences were presented with a visually presented probe word which participants had to decide was related or unrelated to the sentence. The aim of these catch trials was to ensure that attention was paid to the meaning of the sentences. Thus, for each catch sentence, a probe word was selected that was either clearly semantically related (50%) or clearly semantically unrelated (50%) to the sentence’s meaning. Finally, to create a low-level auditory baseline condition, 32 experimental sentences were randomly selected and converted to signal-correlated noise (SCN) using Praat software (<http://www.praat.org>). Conversion to SCN involved replacing all the spectral detail with noise, rendering sentences unintelligible whilst maintaining low-level acoustic properties by retaining the original spectral and amplitude profiles. SCN was chosen as the baseline condition to be able to directly compare these results with those of previous fMRI studies on ambiguity (Rodd et al., 2005, 2012; Bekinschtein et al., 2011). An additional two sentences were selected and converted to SCN for the practice block.

The auditory stimuli were delivered over Sensimetrics insert earphones (<http://www.sens.com/s14/>) in the scanner. EQ

Table 2 | Descriptive statistics [mean(*SD*)] for properties of the ambiguous and unambiguous target words and sentences.

Property		Ambiguous				Unambiguous
		All	Strong biased	Weak biased	Balanced	All
TARGET WORD	N	92	32	27	27	92
	Frequency (log-trans.)	3.61 (1.01)	3.66 (1.16)	3.82 (1.01)	3.24 (0.81)	3.63 (0.93)
	No. letters	4.72 (1.16)	4.50 (1.02)	5.04 (1.43)	4.56 (0.89)	4.76 (1.08)
	No. meanings	1.92 (0.90)	1.81 (0.64)	1.89 (0.97)	2.15 (1.10)	1.09 (0.32)
	No. senses	10.1 (5.60)	9.91 (4.79)	10.4 (7.03)	10.0 (5.06)	4.90 (3.09)
SENTENCE LEVEL	Length (seconds)	2.97 (0.29)	2.96 (0.25)	3.03 (0.30)	2.91 (0.34)	2.97 (0.31)
	No. syllables	16.5 (1.87)	16.3 (1.82)	16.9 (1.92)	16.1 (1.69)	16.4 (1.91)
	No. words	12.5 (1.23)	12.6 (1.34)	12.4 (1.18)	12.6 (1.15)	12.5 (1.23)
	Naturalness rating	5.46 (0.62)	5.37 (0.68)	5.60 (0.61)	5.40 (0.56)	5.80 (0.61)

filtering Software (Sensimetrics, Malden, MA, USA) was used to filter all sound files to ensure accurate frequency reproduction.

DESIGN AND PROCEDURE

An event-related, within-subject design was employed in which participants were presented with all types of sentence trials (ambiguous, unambiguous, SCN and catch sentences) as well as silent (rest) trials. The rest trials were included as another baseline condition, having, on average, the same physical duration as the sentence trials (mean = 3 s; range: 2–4 s). The experiment was divided into four sessions, each with 70 trials: 23 ambiguous; 23 unambiguous; 8 SCN and 8 rest trials as well as two dummy trials to allow for T1 equilibrium before the test trials began. The stimuli were pseudo-randomized so that each run had an equal number of each stimulus type and no ambiguous sentence was placed in the same session as its matched unambiguous sentence in order to avoid potential syntactic priming effects. Each session lasted, on average, 8.47 min. The order of the sessions was counterbalanced across participants.

Each trial commenced with a white fixation cross in the center of a black screen. After 1000 ms, an auditory sentence stimulus or rest trial was presented. Then, for all trials, except catch trials, a silent period of 1500 ms occurred, followed by a jittered inter-trial interval (ITI) of 1000–3000 ms. For catch trials, a silent period of 500 ms followed the sentence offset, then the fixation cross was replaced by a probe word that was presented for 1000 ms on the screen (36 pt bold Helvetica font). Participants indicated whether the probe was related or unrelated to the sentence they just heard by pressing a button with the right index or middle finger. Response button order was counterbalanced across subjects. To discourage participants from actively waiting for a probe to appear and ensure attention to each sentence, we emphasized that responding to the probes would be straightforward if they listened carefully to each sentence. Participants practiced the task inside the scanner before the experimental blocks. The practice block contained a higher proportion of catch-trials than the experimental blocks so that participants could familiarize themselves with the task. A jittered ITI also followed the catch-trials sentences but this ranged from 2000–3000 ms to allow participants at least 3000 ms from probe-onset to respond and prepare for the next trial.

All stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 toolbox (www.vislab.ucl.ac.uk/cogent/index.html). The visual stimuli were projected onto a screen and viewed via mirrors mounted on the head coil. The auditory stimuli were delivered via MRI-compatible insert earphones (Sensimetrics, Malden, MA, USA, Model S-14), which provided a 20–40 dB attenuation level. Participants wore another set of ear protectors over the insert earphones to provide additional attenuation of the scanner noise. The experimenter checked participants could hear the sentences clearly over the noise of the functional EPI sequence prior to the experimental scanning blocks by conducting a practice run in the scanner.

MRI ACQUISITION

Participants were scanned at the Birkbeck-UCL Centre for Neuroimaging (BUCNI) on a Siemens Avanto 1.5T scanner. Whole-brain functional images were acquired with a gradient-echo EPI sequence ($TR = 3000$ ms; $TE = 50$ ms; $3 \times 3 \times 3$ mm resolution). Each run consisted of 180 volumes. In addition, a high-resolution anatomical scan was acquired (T1-weighted FLASH, $TR = 12$ ms; $TE = 5.6$ ms; 1 mm^3 resolution) for anatomical localization purposes.

fMRI DATA ANALYSIS

The functional images were preprocessed and analyzed using Statistical Parametric Mapping software (SPM8, Wellcome Department of Cognitive Neurology, London, UK). Preprocessing involved realignment, spatial normalization and smoothing (8 mm FWHM Gaussian kernel) of the functional images. Entire datasets from three participants were removed because of excessive translational head motion (>3 mm) in at least three of the four scanning sessions. In a further participant, a single scanning session was excluded due to excessive head motion. Finally, for two participants the final five and seven volumes of one run were excluded due to motion. Spatial normalization combined an initial affine component with subsequent non-linear warping (Friston et al., 1995) to best match the Montreal Neurological Institute's MNI-152 template. The resulting images retained their original resolution ($3 \times 3 \times 3$ mm). Two analyses were conducted with separate general linear models. The first model combined all ambiguous

sentences into a single condition regardless of ambiguous word dominance so that parameter estimates of the overall ambiguity effect would not be biased by the differences in sample sizes between the dominance conditions. At the first level, three experimental conditions (ambiguous, unambiguous and SCN) and one “dummy” condition that included the dummy sentences and catch-trials were modeled separately. For each trial, the onset of the sentence/SCN and its duration were specified. For the catch-trials 1.5s was added to the duration to incorporate the presentation of the visual word. Realignment parameters and temporal and dispersion derivatives were included as additional regressors to help model structured noise in the data. The derivatives, in particular, helped accommodate variability in the onset and duration of neural responses to the ambiguous sentences. At the group-level, random effects analyses were employed for two contrasts: “Unambiguous vs. SCN” and “Ambiguous vs. Unambiguous.” The former was conducted first to identify the general language network that is engaged under normal, low-ambiguity, speech. The latter identified the more specific ambiguity-elevated network. For each, the corresponding contrast parameter estimates for each subject were entered into the group-level analysis, where one-sample *t*-tests were computed. Activations were considered significant if they reached a threshold of $p < 0.05$ FWE corrected at the voxel level (Worsley et al., 1996).

The second analysis was identical to the first except that ambiguous sentence trials were modeled as separate dominance conditions. To achieve this, the first-level analysis model included four separate regressors corresponding to the four dominance conditions (strongly biased, weakly biased, balanced, and dominant). For each subject, parameter estimates were obtained for three contrasts: “Strongly biased > Unambiguous,” “Weakly biased > Unambiguous” and “Balanced > Unambiguous.” The dominant condition was not analyzed further due to the small number of trials in this condition. At the group-level, contrast images from these comparisons were entered into a One-Way repeated measures ANOVA to assess effects of dominance across the whole brain and were also employed in region-of-interest (ROI) analyses, described in more detail in the Results section.

Participants’ structural images were normalized to the T1 template and a group mean structural image was created for data display purposes.

RESULTS

BEHAVIORAL RESULTS

On the catch trials participants achieved a mean accuracy of 92% (range = 79–100%), with a mean reaction time of 1328 ms ($SD = 345$), indicating that all participants were paying attention to the meaning of the sentences.

UNAMBIGUOUS SENTENCES vs. SCN

The contrast between unambiguous sentences and the low-level baseline condition showed a significant broad cluster of activation in the left hemisphere centered laterally on the mid-superior temporal sulcus (STS), extending along the length of STS and superiorly to the anterior superior temporal gyrus (STG) (see

Figure 1 and **Table 3**). At a lower threshold ($p < 0.001$ uncorrected), the left anterior temporal activation spread inferiorly into anterior middle temporal cortex. In the right hemisphere, there was a smaller significant cluster of activation centered in mid STG that extended anteriorly into STG and STS. At the lower threshold ($p < 0.001$ uncorrected) it also extended posteriorly and inferiorly into right STS. There was also significant activation in the left dorsolateral precentral gyrus. The LIFG showed activation when the threshold was lowered to $p < 0.001$ uncorrected, specifically within dorsal pars opercularis (peak coordinate $[-54, 17, 19]$; z -score = 3.58). For completeness, the results of the Ambiguous > SCN contrast is presented in the supplementary materials (see Figure S1 and Table S1).

AMBIGUOUS vs. UNAMBIGUOUS SENTENCES

Two clusters in the left hemisphere showed significantly greater activation for ambiguous than unambiguous sentences (see **Figure 2A** and **Table 4**). One cluster was located in the LIFG, centered in pars triangularis. Note that this cluster does not overlap with the pars opercularis cluster that showed greater activation for the unambiguous sentences reported in the previous contrast. At a lower threshold ($p < 0.001$ uncorrected), the activation spread predominately posteriorly through pars opercularis, thereafter extending primarily dorsally in middle frontal/precentral gyrus. The second cluster was located in the posterior inferotemporal cortex (pIT). Its peak was in the posterior occipitotemporal sulcus

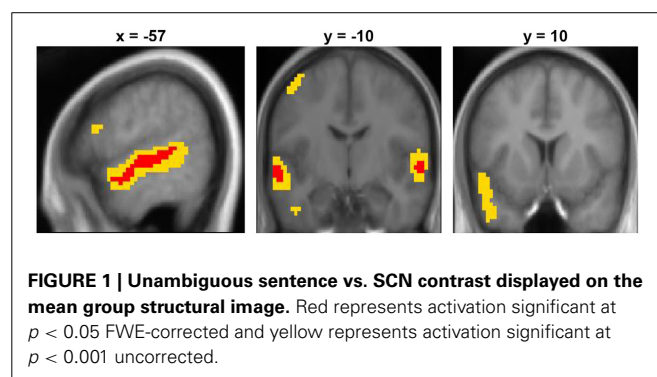


FIGURE 1 | Unambiguous sentence vs. SCN contrast displayed on the mean group structural image. Red represents activation significant at $p < 0.05$ FWE-corrected and yellow represents activation significant at $p < 0.001$ uncorrected.

Table 3 | Unambiguous sentences > SCN: peak activations at $p < 0.05$ FWE corrected.

Brain region	p (corrected)	Z-Score	Co-ordinates (MNI)		
			x	y	z
L STS	<0.001	6.11	-54	-25	-5
L anterior STS	<0.001	5.80	-57	-4	-14
L STS	0.001	5.58	-60	-16	-2
L posterior STS	0.011	5.08	-57	-40	7
R STG	0.001	5.60	60	-10	-2
R anterior STG/STS	0.030	4.88	60	-1	-11
Precentral gyrus	0.022	4.94	-48	-7	58

Sub-peaks that are more than 8 mm from the main peak are indented.

L, left; R, right; STS, superior temporal sulcus, STG, superior temporal gyrus; MTG, Middle temporal gyrus.

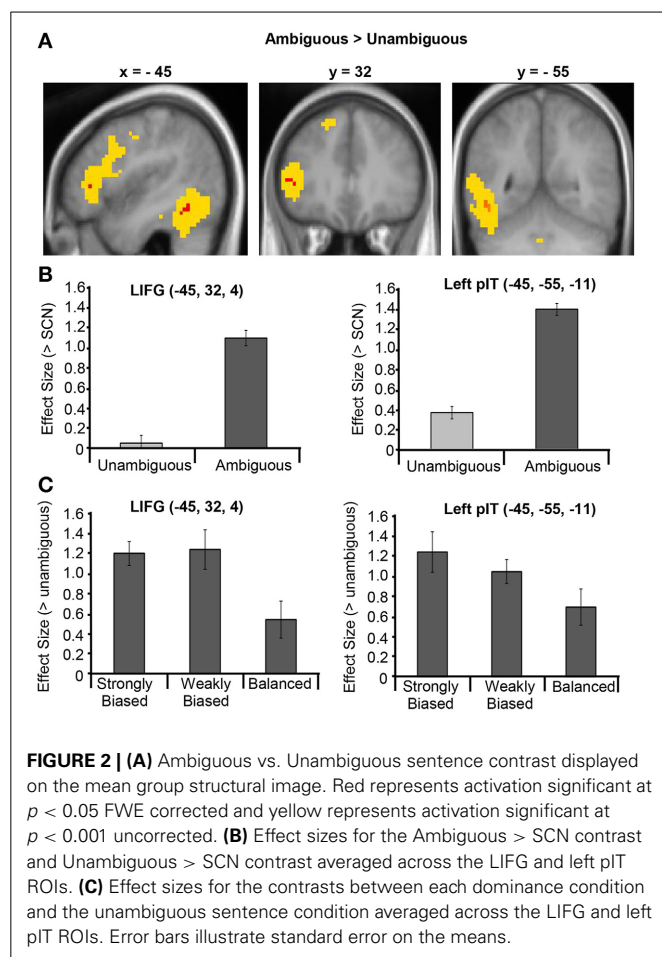


Table 4 | Ambiguous vs. unambiguous sentences: peak activations at $p < 0.05$ FWE corrected.

Brain region	$p(\text{corrected})$	Z-Score	Co-ordinates (MNI)		
			x	y	z
L IFG (pars triangularis)	0.027	4.90	-45	32	4
L OTS	0.011	5.09	-45	-55	-11
L ITG	0.012	5.06	-48	-58	-8

Sub-peaks are indented following main peak.

L, left; OTS, occipitotemporal sulcus; ITG, inferior temporal gyrus; IFG, inferior frontal gyrus.

(OTS) but extended laterally, with a significant sub-peak in the inferior temporal gyrus (LITG). At a lower threshold ($p < 0.001$ uncorrected) this activation extended inferiorly into the posterior and middle portion of the fusiform gyrus, as well as superiorly through the pMTG extending along the STS (see **Figure 2A**).

The response profiles of the regions that showed a significant ambiguity effect were further examined with two region-of-interest (ROI) analyses, performed using the Marsbar toolbox within SPM8 (Brett et al., 2002). The first analysis assessed the nature of the ambiguity difference and the selectivity of these regions' responses to ambiguous sentences by examining

their responses to ambiguous and unambiguous sentences, separately, relative to SCN. For this analysis, a LIFG and a left pIT ROI were constructed as 8 mm radius spheres centered on the LIFG and left pIT group-peak coordinates obtained from the "Ambiguous > Unambiguous" contrast. Mean parameter estimates were obtained in each region for the contrasts "Ambiguous > SCN" and the "Unambiguous > SCN" for each participant. As shown in **Figure 2B**, the ambiguity difference in both regions was, importantly, driven by increased activity for the ambiguous sentences compared to SCN rather than deactivation in the unambiguous condition. In addition, one-sample t -tests revealed that neither the LIFG nor the left pIT ROIs showed a significant response for the unambiguous sentences compared to SCN [$t_{(16)} = 0.17$, $p = 0.87$; $t_{(16)} = 1.88$, $p = 0.25$, respectively].

A second ROI analysis assessed whether these regions were affected by meaning dominance. Mean parameter estimates for the strongly biased, weakly biased and balanced conditions relative to the unambiguous condition were obtained for the LIFG and left pIT ROIs. The resulting effect sizes for each region were normalized relative to the average effect size for that ROI across all participants and all three contrasts. This normalization adjusts for differences in overall effect sizes between ROIs that may confound the magnitude of the differences found between conditions between regions. The normalized effect sizes were entered into a 3×2 repeated-measures ANOVA with Dominance (strongly biased, weakly biased and balanced) and Region (LIFG, pIT) as the two factors. The results showed a significant main effect of Dominance [$F_{(2, 32)} = 3.49$, $p = 0.04$, $\eta_p^2 = 0.18$], no significant main effect of site and no significant Dominance \times Region interaction ($F < 1$ in each case), indicating no reliable differences between the effect of dominance in the frontal and temporal regions. Paired t -tests between each pair of dominance conditions (averaged across region) showed that strongly biased sentences (mean = 1.23, $SD = 0.80$) and weakly biased sentences (mean = 1.15, $SD = 0.77$) produced significantly greater activation than balanced sentences [mean = 0.62, $SD = 0.79$: $t_{(16)} = 2.21$, $p = 0.04$; $t_{(16)} = 2.19$, $p = 0.04$, respectively]. However, there was no significant difference between the strongly and weakly biased sentences [$t_{(16)} = 0.35$, $p = 0.74$]. See **Figure 2C** for the patterns of dominance effects for each of the ROIs.

No significant effects of dominance were found in the whole-brain analysis ($p < 0.05$ FWE corrected).

INTER-SUBJECT VARIABILITY

Although peak co-ordinates from the group analysis identify voxels that show the most reliable effects across subjects, it is also important to assess the inter-subject variability around these peaks. For each subject we obtained the nearest local maximum ($p < 0.05$ uncorrected) to the frontal [-45, 32 4] and temporal group activation peaks [-45, -55, -11] from the Ambiguous > Unambiguous contrast. The location was then examined on each subject's own structural image and identified according to sulcal landmarks. Only peaks that were within the frontal and temporal cortex were considered.

As shown in **Table 5** and **Figure 3**, all subjects, except one, showed significant activation in close proximity to both the frontal and temporal group peaks. Only one subject did not

show any significant activation around the frontal peak, with the nearest local maxima located 28 mm from the peak (coordinates $[-21, 20, 13]$, z -score = 2.85). There was no significant difference between the two group peaks in terms of the average Euclidian distance of the local maxima [paired t -test: $t_{(15)} = 1.37$, $p = 0.19$]. Interestingly, the locations of these local maxima were notably more anatomically consistent (i.e., residing in the same macroanatomic region) in the frontal than in the temporal cortex. For 13 out of the 16 subjects who showed significant activation around the frontal peak, their local maxima resided in pars triangularis, with 2 additional subjects showing activation on the border between pars triangularis (PT_r) and pars orbitalis (PO_r). In contrast, there was more anatomical variability around the temporal peak, with local maxima residing inferiorly within ventral occipital temporal cortex areas, such as OTS and fusiform gyrus (FSG), whilst others were located more laterally within MTG/ITG.

DISCUSSION

The results of this study replicate previous findings of increased activation in the LIFG and posterior temporal cortex for (temporarily) semantically ambiguous sentences compared to unambiguous sentences (Rodd et al., 2005, 2010b, 2012; Davis et al., 2007; Mason and Just, 2007; Zempleni et al., 2007; Bekinschtein et al., 2011). The current study employed ambiguous sentences for which listeners were likely to initially select the incorrect meaning of the ambiguous word and then need to reinterpret their understanding of the sentence later in the comprehension process. This was achieved by presenting the disambiguating information several words after the ambiguous word (e.g., “the woman had to make the TOAST with a very old *microphone*”),

as various psycholinguistic models of ambiguity resolution claim that listeners make an initial meaning selection within a few hundred milliseconds of hearing an ambiguous word (Twilley and Dixon, 2000). Thus, this initial finding of ambiguity-responsive activation in the LIFG and posterior temporal cortex is consistent with the hypothesis that both of these regions may be important for reinterpreting the meaning of a word during sentence comprehension (e.g., Novick et al., 2005; Zempleni et al., 2007; Rodd et al., 2012).

This study, furthermore, explored the roles of these regions in ambiguity resolution by assessing their response profiles to different types of sentence stimuli as well as the inter-subject consistency of these regions' responses to ambiguity. The results of the functional-based analyses are discussed first, separately for the two regions, followed by discussion of the inter-subject variability.

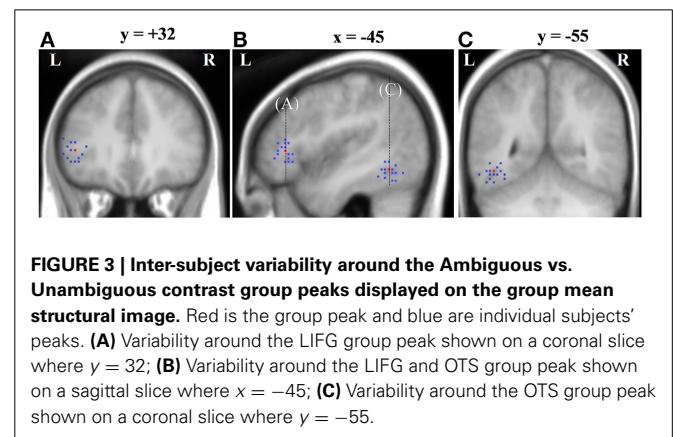


Table 5 | Individual subjects' "Ambiguous > Unambiguous" local maxima nearest to the frontal and temporal group peaks.

Subject	LIFG $[-45\ 32\ 4]$						L OTS $[-45\ -55\ -11]$					
	x	y	z	Z score	Distance from peak	Region	x	y	z	Z score	Distance from peak	Region
1	-36	35	7	2.20	9.9	IFS	-39	-52	-5	3.24	9.0	FSG
2	-45	32	10	2.94	6.0	PT _r	-45	-49	-11	2.56	6.0	OTS
3	-45	29	-5	1.87	9.5	PT _r	-48	-55	-5	2.57	6.7	MTG
4	-48	38	4	4.75	6.7	PT _r	-48	-58	-11	2.06	4.2	ITG/OTS
5	-48	38	10	2.32	9.0	PT _r	-45	-52	-14	2.53	4.2	OTS
6	-48	32	-5	2.62	9.5	PO _r /PT _r	-45	-58	-20	3.53	9.5	ITG
7	-48	26	-5	2.54	11.2	PT _r	-42	-67	-14	2.39	12.7	ITG/FSG
8	-39	38	1	4.26	9.0	PT _r	-39	-52	-11	2.61	6.7	OTS/ITG
9	-54	29	1	2.53	9.9	PT _r	-48	-52	-14	1.81	5.2	MTG/ITG
10			n/a				-36	-52	-17	1.76	11.2	FSG
11	-51	29	-2	2.84	9.0	PT _r	-57	-61	-14	3.31	13.7	MTG
12	-48	29	10	2.52	7.3	PT _r	-51	-61	-20	2.83	12.4	ITG
13	-54	32	13	1.87	12.7	PT _r	-48	-46	-17	2.42	11.2	OTS
14	-54	26	1	3.62	11.2	PT _r	-48	-58	-11	2.13	4.2	MTG
15	-51	32	7	2.07	6.7	PT _r	-36	-58	-14	3.11	9.9	FSG
16	-42	41	-2	2.02	11.2	PO _r /PT _r	-51	-58	-11	3.63	6.7	MTG
17	-42	29	-5	1.79	9.9	PT _r	-42	-61	-8	1.82	7.3	FSG
Mean	-47	32	3	2.67	9.3 mm		-45	-56	-13	2.60	8.3 mm	

Two specific functional questions were assessed. (1) Is activation within these regions specific to ambiguous sentences or present for all sentences, albeit to a less extent for low-ambiguity sentences? (2) Are these regions primarily contributing to semantic reinterpretation processes or initial meaning selection components of ambiguity resolution? For these questions, two contrasts were assessed via ROIs around the frontal and temporal group peak co-ordinate separately: (1) the regions' response to unambiguous sentences compared to a low-level auditory baseline and (2) the modulation of these responses by meaning dominance (i.e., meaning frequency) by comparing biased and balanced ambiguous words. A region showing an ambiguity effect that is primarily involved in semantic reinterpretation will show larger responses for biased than balanced sentences, whereas regions that are primarily involved in initial meaning selection will show the reverse profile. Together the results of these two contrasts give insights into the ways by which these regions support ambiguity resolution and language comprehension more generally, which ultimately help constrain theories of their functions in these processes.

LEFT INFERIOR FRONTAL GYRUS

Statistically robust activation ($p < 0.05$ FWE corrected) for semantically ambiguous sentences was found in the middle portion of the LIFG, namely pars triangularis (**Figure 2A**). This region has been reported in nearly all published studies on semantically ambiguous sentences (Rodd et al., 2005, 2010b, 2012; Davis et al., 2007; Mason and Just, 2007; Zempleni et al., 2007; Bekinschtein et al., 2011). Thus, this study corroborates it as the most consistent site of significant ambiguity-elevated peaks in the frontal cortex.

The results of the two additional contrasts showed two important findings pertaining to the role of this region in language comprehension. First, this region showed no significant response to unambiguous sentences compared to SCN (**Figure 2B**), suggesting that it may not be routinely involved during comprehension of low-ambiguity speech and may, therefore, perform different functions to those involved in general sentence processing. Several other neuroimaging studies have also failed to find significant LIFG responses to low-ambiguity sentences (Crinion et al., 2003; Spitsyna et al., 2006; Rodd et al., 2012).

This response selectivity for ambiguous but not unambiguous sentences is most consistent with the conflict resolution account of LIFG function (Thompson-Schill et al., 1997; Novick et al., 2005, 2009). According to this theory, the LIFG is involved in sentence comprehension only when there is conflict between simultaneously active representations in order to support the selection of one alternative. It is worth noting that although this region is not recruited by the relatively simple low-ambiguity sentences used in this study, its role is very unlikely to be specific to resolving semantic ambiguity as activation in this region has been observed for a range of other types of complex sentences including syntactically ambiguous sentences and syntactically complex sentences (e.g., Santi and Grodzinsky, 2010; Tyler et al., 2011).

The lack of a response for unambiguous sentences is less easily compatible with sentence-general accounts of the LIFG. For

example, Hagoort's unification theory (Hagoort, 2005, 2013) proposes that the LIFG serves to combine small units of linguistic information into larger representations of a sentence. Therefore, all sentences should engage this region to some extent. Although the lack of an unambiguous response may merely be masked by activation in the baseline condition (Binder et al., 1999), patient data provide some corroborating evidence that the LIFG may not be necessary for and, thus, not always involved in language comprehension. For example, patients with LIFG lesions have relatively preserved comprehension of words and of relatively simple sentences (Caramazza and Zurif, 1976; Caplan et al., 1996; Yee et al., 2008; Novick et al., 2009).

It is important to note that the whole LIFG was not uniform in its response to the unambiguous condition. A more posterior region, in pars opercularis, showed greater activation for unambiguous sentences as well as an additional response to the ambiguous stimuli, although both of these effects were only significant at a more lenient statistical threshold. Thus, this suggests that there may be functionally distinct regions in the LIFG, some of which perform processes that are general to sentences and others that are more specific to certain types of sentences. However, it is not clear how this can be reconciled with claims that the function of the LIFG can be fractionated on the basis of either the linguistic nature of the processes (Gough et al., 2005; Vigneau et al., 2006; Hagoort, 2013) or the nature of the cognitive operation (Novick et al., 2005; Badre and Wagner, 2007).

The second key question concerned the effect of dominance (i.e., meaning frequency). The results revealed that mid-LIFG activation was greater for ambiguous sentences that contained a biased ambiguous word, which have one particularly dominant meaning (e.g., "toast"), than a balanced ambiguous word whose meanings are relatively equally frequent (e.g., "bark"; **Figure 2C**). This finding further supports the reinterpretation hypothesis, as listeners are more likely to reinterpret the meaning of a biased word because they were always disambiguated to their subordinate meaning (e.g., speech meaning of "toast"). Psycholinguistic research demonstrates that listeners and readers usually initially select the dominant meaning of a biased word when encountered before disambiguating context (e.g., the bread meaning of "toast"), whereas for balanced words there is less systematic bias for either alternative meaning across individuals (e.g., some may select the dog meaning of "bark" while others select the tree meaning). Thus, for biased sentences, the initial interpretation would often be incorrect and, hence, need to be reinterpreted more often than for balanced sentences. Although no significant dominance effects were found in the whole-brain voxel-wise analysis, this may reflect the fact that dominance responses are likely to be highly variable across both voxels and subjects, given the findings that meaning preferences are inherently variable across subjects (Rodd et al., 2013) and that the exact time-course of disambiguation varies across sentences (Rodd et al., 2012) and individuals depending on comprehension ability (Gernsbacher et al., 1990; Gernsbacher and Robertson, 1995).

The results of the dominance contrast directly replicate Mason and Just's (2007) finding of greater LIFG activation for biased than balanced sentences in visually-presented sentences, albeit in a more anterior ventral region, and provide the first evidence of

these effects in spoken sentences. The results also converge with other dominance effects found in the LIFG, including greater activation for subordinate-biased compared with dominant-biased sentences (Zemleni et al., 2007) and the finding of a negative correlation between LIFG activation and the dominance of syntactically ambiguous sentences (Rodd et al., 2010b). Again, both of these effects reflect greater activation for sentences where reinterpretation is more likely. The finding that similar dominance effects in this region were found for this set of spoken sentences as has been reported for visually-presented sentences (Mason and Just, 2007; Zemleni et al., 2007) suggests a common system for disambiguating spoken and written sentences.

Interestingly, the current results showed that the two types of biased sentences patterned together: activation for strongly- and weakly-biased sentences was significantly greater than for balanced sentences but were not significantly different from each other. This suggests a non-linear relationship between dominance and neural response with neural responses not simply being associated with the likelihood of semantic reinterpretation. One possible reason for this pattern is that the neural responses may, in part, reflect how *difficult* reinterpretation is because this process is more demanding for biased than balanced words regardless of the extent of the bias *per se*. This explanation is derived from a large body of psycholinguistic research that demonstrates a difference in the state of the alternative meanings of biased and balanced words during the comprehension of late-disambiguation sentences. When biased words are encountered before disambiguating context, their dominant meaning is quickly integrated (e.g., the bread meaning of toast) while their subordinate (speech) meanings are quickly suppressed or not accessed at all. In contrast, multiple meanings of balanced words (tree and dog meanings of “bark”) are initially activated and it takes longer for one meaning to be integrated (e.g., Simpson, 1994; Twilley and Dixon, 2000; Duffy et al., 2001). Thus, contextually appropriate, subordinate, meanings may be harder to (re)integrate than non-selected balanced meanings because they are less available when the disambiguating information is later encountered and contextually-inappropriate, dominant, meanings may also be harder to override than initially-selected balanced meanings because they have been more strongly integrated (Simpson, 1994; Twilley and Dixon, 2000; Duffy et al., 2001; Gernsbacher and St John, 2001). Thus, the dominance pattern suggests that the LIFG may be particularly important to integrate less available meanings and/or suppress dominant incorrect representations during sentence reinterpretation. This is highly consistent with a recent patient study demonstrating that patients with damage to the LIFG had particular difficulty in resolving subordinate-biased sentences compared to sentences with balanced ambiguous words (Vuong and Martin, 2011). It is also compatible with Novick et al.’s (2005) view that the LIFG’s role in sentence comprehension is to resolve misanalyses, although they refer to syntactic misinterpretations, and converges with findings in non-linguistic domains, such as emotion regulation where two recent meta-analyses have shown that the LIFG (and posterior temporal cortex) is engaged during reinterpretation of emotionally eliciting events (Buhle et al., 2013; Kohn et al., 2014).

This reinterpretation-based conclusion is predicated on the assumption that the greater activation for biased words is related to processes occurring at the time of the disambiguating information. It is therefore important to rule out alternative explanations that could potentially account for these effects in terms of processing at the time that they are initially encountered. At face value, such accounts seem unlikely as no current cognitive theories predict that there should be greater cognitive processing when encountering ambiguous words with one strongly dominant meaning compared with balanced words with two equally-frequent meanings, when these words occur in a neutral context. While it is, in theory, possible that biased words could induce greater processing demands if participants had learnt during the course of the experiment that when they encountered an ambiguous word they should interpret it with the less preferred meaning, existing behavioral and neuroimaging research strongly suggest that such expectations are either not learnt, or do not substantially influence, sentence comprehension. For example, numerous behavioral studies that have examined processing of these late-disambiguation sentences show that listeners’ and readers do not experience behavioral processing costs (i.e., longer reading times or poorer performance on a secondary concurrent task) when they encounter biased ambiguous words in a sentence but only experience processing costs when the disambiguating information is encountered later in the sentence (e.g., Duffy et al., 1988, 2001; Rodd et al., 2010a). If biased words induced greater selection conflict at initial encounter with the word then such costs should be found at the time of the ambiguous word. In addition, even in fMRI studies where the ambiguity is concealed, such that participants do not report noticing any ambiguity in the sentences, activation is found in broadly similar brain regions, suggesting that such activity does not reflect greater selection conflict arising from an explicit strategy employed by the listener (Rodd et al., 2005).

In contrast to the current results, which emphasize the role of this region in reinterpretation, various theories suggest that the LIFG should also be important for processes associated with initial meaning selection (in the absence of reinterpretation) whenever this induces conflict (Thompson-Schill et al., 1997; Novick et al., 2009) and/or makes unification difficult (Hagoort, 2013). In addition, Rodd et al. (2012) found evidence that this region responds to both reinterpretation and initial meaning selection stages of ambiguity processing. Although the current results showed greater activation for sentences with a higher likelihood of reinterpretation, the results cannot rule out the possibility that it also responds to initial selection demands but the fMRI protocol was not sensitive enough to detect them. Future research would benefit from using techniques with higher temporal resolution than the fMRI protocol used here, as these processes occur at different times during sentence processing, such as magnetoencephalography (MEG) or time-sensitive fMRI techniques (Rodd et al., 2012), and should compare both the existence and magnitude of these responses.

In summary, the results replicate the involvement of the LIFG in ambiguity resolution and additionally show that this

ambiguity-responsive region of the LIFG is not significantly engaged by all types of sentences to the same extent. This region shows no significant response to unambiguous sentences and demonstrates a larger ambiguity response for ambiguous sentences that are more likely to require reinterpretation. Together, the results are most consistent with accounts of this region that do not view LIFG as mandatory for language comprehension (e.g., conflict resolution account) and suggests that it supports comprehension when the listener's current interpretation needs to be updated in light of new contextual information.

POSTERIOR TEMPORAL CORTEX

In the temporal lobe, statistically robust activation for the semantically ambiguous sentences was located in the left posterior inferior temporal cortex (pIT), specifically in the occipitotemporal sulcus and inferior temporal gyrus. This is in a similar location to that found by Rodd et al. (2012) and Bekinschtein et al. (2011), but is more inferior than other studies where activation centers around pMTG/ITG (Rodd et al., 2005; Davis et al., 2007; Zemleni et al., 2007).

The results of the subsequent experimental contrasts showed that this region had a highly similar response profile to the mid-LIFG. The analyses showed (1) no significant response to unambiguous sentences (**Figure 2B**) and (2) the same pattern of dominance effects, where activation was greater for biased than balanced sentences (**Figure 2C**). Together, the results suggest that this region of the pIT is also involved in semantic reinterpretation processes which are not required for comprehension of low-ambiguity sentences.

The locus of this activation is interesting because it is posterior to regions more strongly associated with multimodal semantic processing, namely the anterior fusiform gyrus (Binder et al., 2009; Price and Devlin, 2011; Seghier and Price, 2011), and the cluster is more inferior than that associated with other lexical/semantic processes such as sound-to-meaning mapping in the pMTG/ITS (Hickok and Poeppel, 2007) and semantic control in the pMTG (Jefferies, 2013). Instead, this region has been more generally attributed to high-level visual processing associated with either the visual form of words (Dehaene and Cohen, 2011) or with visual features of meaningful stimuli more generally (Martin, 2007; Price and Devlin, 2011). This region is not consistently found in auditory single word or spoken sentence studies (Binder et al., 2000; Xiao et al., 2005; Spitsyna et al., 2006; Davis and Gaskell, 2009; Obleser and Kotz, 2010), but a large body of research shows that the response of this region is strongly modulated by non-visual processes such as semantics and phonological information (Devlin et al., 2006; Song et al., 2010; Yoncheva et al., 2010; Twomey et al., 2011) and can be activated in the absence of visual information (e.g., Mellet et al., 1998; Price et al., 2003). Thus, activation in response to ambiguity may reflect top-down accessing of visual information related to orthographic representations and/or visual attributes of the objects referred to in the sentence.

This view makes no strong prediction about whether this response should also occur for unambiguous sentences since these kinds of sentences may also evoke visual information.

Indeed a recent study has reported activation for low-ambiguity speech in this region (Rodd et al., 2012). The lack of a response to unambiguous sentences, however, is incompatible with accounts that claim that such visual activation is a fundamental component of semantic processing (Martin, 2007).

Ambiguous sentences may engage visual information processing in various ways. For example, the ambiguity may evoke a visual image of the ambiguous word or an image of the content of the ambiguous sentence, which is supported by a large body of research showing increased activation of visual processing areas during imagery tasks (D'Esposito et al., 1997; Mellet et al., 1998; Martin, 2007; Dehaene and Cohen, 2011). Alternatively, visual representations may be activated more automatically by the increased level of semantic competition induced by the ambiguous words (Gennari et al., 2007), given the evidence of inherent functional and anatomical connections between semantic and perceptual representations (Kherif et al., 2011; Price and Devlin, 2011).

While the locus of this temporal activation is most consistent with regions discussed in visual processing accounts, it must be emphasized that it is also close to regions imputed in other accounts of posterior temporal function. In particular, this region is just inferior to pMTG/ITS that is argued to support sound-meaning mapping (Hickok and Poeppel, 2007; Hickok, 2012). Thus, the finding of an ambiguity effect in the broad vicinity of this region may also be considered consistent with this account, as the mapping between sound and meaning is more uncertain for ambiguous than unambiguous words. Presumably, this mapping needs re-computing when the meaning of word is not supported by contextual information (Rodd et al., 2012), which is further supported by the finding that this region was affected by reinterpretation load. However, it is difficult to explain the lack of significant response for unambiguous sentences in this region if it supports such a fundamental aspect of speech comprehension. Instead, the temporal areas that showed responses to unambiguous sentences were located more superiorly, along the STG/STS and anterior MTG. This distribution is consistent with previous studies on speech comprehension, where activation for low-ambiguity speech is typically confined to superior/middle temporal cortex (Humphries et al., 2001; Spitsyna et al., 2006; Adank and Devlin, 2010; Obleser and Kotz, 2010) rather than extending into inferior temporal regions in the way that is typically seen for studies of ambiguity resolution (Rodd et al., 2005; Davis et al., 2007; Zemleni et al., 2007; Bekinschtein et al., 2011).

In summary, the results replicate the involvement of the posterior inferior temporal cortex in ambiguity resolution and further show that activation in this region is not present for low-ambiguity sentences and is particularly responsive to ambiguous sentences that require reinterpretation. Like the response of the mid-LIFG, the results are most consistent with accounts of this region that impute functions that are not mandatory for sentence processing (e.g., visual-based processes) and suggests that this region also supports comprehension particularly when listeners need to update their understanding of a sentence in light of new contextual information.

INTER-SUBJECT VARIABILITY

As this study confirms, the involvement of both frontal and temporal regions in the processing of semantically ambiguous sentences is emerging as a highly consistent finding across fMRI studies. However, these results are based on group-level analyses, which do not indicate the extent to which this reflects a network in which all components are engaged by all subjects. To investigate this, inter-subject variability was assessed in relation to the frontal and temporal group peaks. All subjects, except one, showed ambiguity-related local maxima within 10 mm of the LIFG and posterior temporal cortex group peak. These findings provide evidence that the group-level results reflect activation patterns that are consistent across a majority of subjects, rather than being driven by large activations in only a small proportion of individuals.

Other interesting findings also came out of this analysis. First, the anatomical locations of the LIFG individual peaks were highly consistent, being located within pars triangularis in over 80% of subjects. This further highlights the potential importance of this particular LIFG sub-division in semantic ambiguity resolution. In contrast, the locations of the temporal peaks were more anatomically variable. While the majority of subjects showed peaks in inferior, as oppose to middle, temporal regions (ITG, occipitotemporal sulcus, fusiform gyrus vs. MTG), there was no clearly consistent anatomical field. Such anatomical inconsistency in the temporal cortex's response to ambiguity across participants might explain why different studies have reported activation in these different sub-regions. The nature of this inter-subject variability is currently unclear, although several possible explanations exist. MRI and post-mortem investigations of the morphology of the temporal lobe have found that various macroanatomical structures (e.g., the inferior temporal sulcus) are extremely variable across subjects (Ono et al., 1990; Kim et al., 2000). The posterior inferior temporal cortex in particular has also been observed to have less distinct cytoarchitectonic boundaries such that neurologists have reported difficulty in subdividing this region based on microcellular properties (von Economo, 2009). These findings suggest that the relationship between function and macroanatomically-defined regions might be less consistent in the region and, thus, across subjects. Recent fMRI has further shown evidence that higher-level cognitive processes, more generally, show looser function-anatomy mappings than lower-level cognitive processes (Duncan et al., 2009; Tahmasebi et al., 2012). Alternatively, it is possible that this inter-subject variability found in this study may reflect functionally different responses to the ambiguous sentences across subjects, such that subjects draw on different cognitive operations to resolve the ambiguity. Although the reasons for such variability are currently uncertain, these findings clearly show inter-subject consistency of both frontal and temporal regions in processing ambiguous sentences.

ADDITIONAL AMBIGUITY-RESPONSIVE REGIONS

Inspecting the data at a lower statistical threshold revealed that ambiguity-elevated activations occurred across substantially larger clusters within the frontal and temporal cortex than that shown when applying stringent statistical threshold. The frontal cluster extended throughout pars triangularis and pars

opercularis. However, interestingly, activation was not found in its most anterior sub-division, pars orbitalis. This is particularly noteworthy as anterior LIFG has been specifically attributed to semantic processing (Poldrack et al., 1999; Gough et al., 2005; Hagoort, 2005, 2013; Vigneau et al., 2006; Badre and Wagner, 2007). This result is not entirely unexpected as the response of anterior LIFG to semantically ambiguous sentences is the least consistent of the three sub-divisions, with only two studies reporting activation across all three sub-divisions (Rodd et al., 2010b, 2012). One potential explanation is that this region serves a specific semantic-related function that is not important for resolving all types of ambiguous sentences. For example, one current theory of the anterior LIFG is that it supports controlled semantic retrieval (Badre and Wagner, 2007). In these sentences, the disambiguating word may have constituted a sufficiently strong semantic cue to the correct meaning of the word that additional retrieval processes were not needed.

Another interesting observation was the notable extension of ambiguity-related activation into frontal and temporal regions that have been strongly implicated in phonological processing, namely the posterior and mid-STS as well as the posterior LIFG (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Hagoort, 2013). Such activation may reflect a replaying of the heard sentence in an attempt to reanalyze the meaning of these sentences. Thus, these additional results may provide working hypotheses for both cognitive and neural models of ambiguity resolution. It is, also, possible that these less robust regions may reflect inter-subject variability in the processing of ambiguous sentences.

Together these findings replicate the involvement of the LIFG and posterior temporal cortex in semantic ambiguity resolution found in previous studies and further demonstrate that this network is highly consistent across individuals. The results, additionally, explored the potential roles of these regions in this process, supporting the hypothesis that both regions may be particularly important when listeners need to reinterpret the meaning of an ambiguous word during sentence comprehension.

ACKNOWLEDGMENTS

This work was funded by a BBSRC studentship awarded to Sylvia Vitello and a Leverhulme Trust grant awarded to Jennifer M. Rodd and Joseph T. Devlin. We would like to thank Pamela Farago, Alice Treen and Sara Watchko for help with data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2014.00530/abstract>

REFERENCES

- Adank, P., and Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *Neuroimage* 49, 1124–1132. doi: 10.1016/j.neuroimage.2009.07.032
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Badre, D., and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45, 2883–2901. doi: 10.1016/j.neuropsychologia.2007.06.015

- Bekinschtein, T. A., Davis, M. H., Rodd, J. M., and Owen, A. M. (2011). Why clowns taste funny: the relationship between humor and semantic ambiguity. *J. Neurosci.* 31, 9665–9671. doi: 10.1523/JNEUROSCI.5058-10.2011
- Binder, J., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055
- Binder, J., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi: 10.1093/cercor/10.5.512
- Binder, J., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., and Cox, R. W. (1999). Conceptual processing during the conscious resting state. A functional MRI study. *J. Cogn. Neurosci.* 11, 80–95. doi: 10.1162/089892999.563265
- Brett, M., Anton, J.-L., Valabregue, R., and Poline, J.-B. (2002). “Region of interest analysis using an SPM toolbox [abstract],” in *Presented at the 8th International Conference on Functional Mapping of the Human Brain*. Vol. 16 (Sendai). Available on CD-ROM in NeuroImage, Abstract 497.
- Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., et al. (2013). Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cereb. Cortex*. doi: 10.1093/cercor/bht154. [Epub ahead of print].
- Caplan, D., Hildebrandt, N., and Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain* 119, 933–949. doi: 10.1093/brain/119.3.933
- Caramazza, A., and Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension - evidence from Aphasia. *Brain Lang.* 3, 572–582. doi: 10.1016/0093-934X(76)90048-1
- Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., and Wise, R. J. S. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain* 126, 1193–1201. doi: 10.1093/brain/awg104
- Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16032–16037. doi: 10.1073/pnas.0701309104
- Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111
- Dehaene, S., and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262. doi: 10.1016/j.tics.2011.04.003
- D’Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., et al. (1997). A functional MRI study of mental image generation. *Neuropsychologia* 35, 725–730. doi: 10.1016/S0028-3932(96)00121-2
- Devlin, J. T., Jamison, H. L., Gonnerman, L. M., and Matthews, P. M. (2006). The role of the posterior fusiform gyrus in reading. *J. Cogn. Neurosci.* 18, 911–922. doi: 10.1162/jocn.2006.18.6.911
- Duffy, S. A., Kambe, G., and Rayner, K. (2001). “The effect of prior disambiguating context on the comprehension of ambiguous words: evidence from eye movements,” in *On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity*, ed D. S. Gorfein (Washington, DC: American Psychological Association), 27–43.
- Duffy, S. A., Morris, R. K., and Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *J. Mem. Lang.* 27, 429–446. doi: 10.1016/0749-596X(88)90066-6
- Duncan, K. J., Pattamadilok, C., Knierim, I., and Devlin, J. T. (2009). Consistency and variability in functional localisers. *Neuroimage* 46, 1018–1026. doi: 10.1016/j.neuroimage.2009.03.014
- Friston, K., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., and Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Hum. Brain Mapp.* 3, 165–189. doi: 10.1002/hbm.460030303
- Gennari, S. P., MacDonald, M. C., Postle, B. R., and Seidenberg, M. S. (2007). Context-dependent interpretation of words: Evidence for interactive neural processes. *Neuroimage* 35, 1278–1286. doi: 10.1016/j.neuroimage.2007.01.015
- Gernsbacher, M. A. (1991). “Cognitive processes and mechanisms in language comprehension: the structure building framework,” in *The Psychology of Learning and Motivation*, ed G. H. Bower (New York, NY: Academic Press), 217–263.
- Gernsbacher, M. A., and Robertson, R. R. W. (1995). Reading skill and suppression revisited. *Psychol. Sci.* 165–169. doi: 10.1111/j.1467-9280.1995.tb00326.x
- Gernsbacher, M. A., and St John, M. F. (2001). “Modeling suppression in lexical access,” in *On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity*, ed D. S. Gorfein (Washington, DC: American Psychological Association), 47–65.
- Gernsbacher, M. A., Varner, K. R., and Faust, M. E. (1990). Investigating differences in general comprehension skill. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 430–445. doi: 10.1037/0278-7393.16.3.430
- Gough, P. M., Nobre, A. C., and Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci.* 25, 8010–8016. doi: 10.1523/JNEUROSCI.2307-05.2005
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004
- Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Front. Psychol.* 4:416. doi: 10.3389/fpsyg.2013.00416
- Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *J. Commun. Disord.* 45, 393–402. doi: 10.1016/j.jcomdis.2012.06.004
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Humphries, C., Willard, K., Buchsbaum, B., and Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport* 12, 1749–1752. doi: 10.1097/00001756-200106130-00046
- Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex* 49, 611–625. doi: 10.1016/j.cortex.2012.10.008
- Kherif, F., Josse, G., and Price, C. J. (2011). Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cereb. Cortex* 21, 103–114. doi: 10.1093/cercor/bhq063
- Kim, J.-J., Crespo-Facorro, B., Andreasen, N. C., O’Leary, D. S., Zhang, B., Harris, G., et al. (2000). An MRI-based parcellation method for the temporal lobe. *Neuroimage* 11, 271–288. doi: 10.1006/nimg.2000.0543
- Kohn, N., Eickhoff, S., Scheller, M., Laird, A., Fox, P., and Habel, U. (2014). Neural network of cognitive emotion regulation—an ALE meta-analysis and MACM analysis. *Neuroimage* 87, 345–355. doi: 10.1016/j.neuroimage.2013.11.001
- Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45. doi: 10.1146/annurev.psych.57.102904.190143
- Mason, R. A., and Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Res.* 1146, 115–127. doi: 10.1016/j.brainres.2007.02.076
- Mellet, E., Tzourio, N., Denis, M., and Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport* 9, 803–808. doi: 10.1097/00001756-199803300-00007
- Novick, J. M., Kan, I. P., Trueswell, J. C., and Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cogn. Neuropsychol.* 26, 527–567. doi: 10.1080/02643290903519367
- Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). Cognitive control and parsing: reexamining the role of Broca’s area in sentence comprehension. *Cogn. Affect. Behav. Neurosci.* 5, 263–281. doi: 10.3758/CABN.5.3.263
- Obleser, J., and Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex* 20, 633–640. doi: 10.1093/cercor/bhp128
- Ono, M., Kubik, S., and Abernathy, C. D. (1990). *Atlas of the Cerebral Sulci*. Stuttgart: Thieme.
- Parks, R., Ray, J., and Bland, S. (eds.). (1998). *Wordsmyth English dictionary-Thesaurus*. Chicago, IL: University of Chicago.
- Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10, 15–35. doi: 10.1006/nimg.1999.0441
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847. doi: 10.1016/j.neuroimage.2012.04.062
- Price, C. J., and Devlin, J. T. (2011). The Interactive Account of ventral occipito-temporal contributions to reading. *Trends Cogn. Sci.* 15, 246–253. doi: 10.1016/j.tics.2011.04.001
- Price, C. J., Winterburn, D., Giraud, A. L., Moore, C. J., and Noppeney, U. (2003). Cortical localisation of the visual and auditory word form areas: a

- reconsideration of the evidence. *Brain Lang.* 86, 272–286. doi: 10.1016/S0093-934X(02)00544-8
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- Rayner, K., and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Mem. Cognit.* 14, 191–201. doi: 10.3758/BF03197692
- Rodd, J. M., Davis, M. H., and Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb. Cortex* 15, 1261–1269. doi: 10.1093/cercor/bhi009
- Rodd, J. M., Gaskell, M. G., and Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *J. Mem. Lang.* 46, 245–266. doi: 10.1006/jmla.2001.2810
- Rodd, J. M., Johnsrude, I. S., and Davis, M. H. (2010a). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain Lang.* 115, 182–188. doi: 10.1016/j.bandl.2010.07.005
- Rodd, J. M., Johnsrude, I. S., and Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cereb. Cortex* 22, 1761–1773. doi: 10.1093/cercor/bhr252
- Rodd, J. M., Longe, O. A., Randall, B., and Tyler, L. K. (2010b). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia* 48, 1324–1335. doi: 10.1016/j.neuropsychologia.2009.12.035
- Rodd, J. M., Lopez Cutrin, B., Kirsch, H., Millar, A., and Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *J. Mem. Lang.* 68, 180–198. doi: 10.1016/j.jml.2012.08.002
- Santi, A., and Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage* 51, 1285–1293. doi: 10.1016/j.neuroimage.2010.03.034
- Seghier, M. L., and Price, C. J. (2011). Explaining left lateralization for words in the ventral occipitotemporal cortex. *J. Neurosci.* 31, 14745–14753. doi: 10.1523/JNEUROSCI.2238-11.2011
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., and Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cogn. Psychol.* 14, 489–537. doi: 10.1016/0010-0285(82)90017-2
- Seidenberg, M. S., Waters, G. S., Sanders, M., and Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Mem. Cognit.* 12, 315–328. doi: 10.3758/BF03198291
- Sereno, S. C. (1995). Resolution of lexical ambiguity: evidence from an eye movement priming paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 582–595. doi: 10.1037/0278-7393.21.3.582
- Simpson, G. B. (1994). Context and the processing of ambiguous words. *Handbook Psycholinguist.* 22, 359–374.
- Simpson, G. B., and Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *J. Mem. Lang.* 30, 627–643. doi: 10.1016/0749-596X(91)90029-J
- Song, Y., Bu, Y., Hu, S., Luo, Y., and Liu, J. (2010). Short-term language experience shapes the plasticity of the visual word form area. *Brain Res.* 1316, 83–91. doi: 10.1016/j.brainres.2009.11.086
- Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., and Wise, R. J. (2006). Converging language streams in the human temporal lobe. *J. Neurosci.* 26, 7328–7336. doi: 10.1523/JNEUROSCI.0559-06.2006
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *J. Verbal Learn. Verbal Behav.* 18, 645–659. doi: 10.1016/S0022-5371(79)90355-4
- Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., et al. (2012). Is the link between anatomical structure and function equally strong at all cognitive levels of processing? *Cereb. Cortex* 22, 1593–1603. doi: 10.1093/cercor/bhr205
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14792–14797. doi: 10.1073/pnas.94.26.14792
- Twilley, L. C., and Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychon. Bull. Rev.* 7, 49–82. doi: 10.3758/BF03210725
- Twilley, L. C., Dixon, P., Taylor, D., and Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Mem. Cogn.* 22, 111–126. doi: 10.3758/BF03202766
- Twomey, T., Kawabata Duncan, K. J., Price, C. J., and Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *Neuroimage* 55, 1242–1251. doi: 10.1016/j.neuroimage.2011.01.001
- Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., et al. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain* 134, 415–431. doi: 10.1093/brain/awq369
- Vigneau, M., Beaucousin, V., Herve, P.-Y., Duffau, H., Crivello, F., Houde, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432. doi: 10.1016/j.neuroimage.2005.11.002
- Vitello, S. (2014). *Cognitive and Neural Mechanisms Underlying Semantic Ambiguity Resolution*. Unpublished doctoral dissertation, University College London.
- von Economo, C. (2009). *Cellular Structure of the Human Cerebral Cortex*. Basel: Karger Medical and Scientific Publishers.
- Vuong, L. C., and Martin, R. C. (2011). LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain Lang.* 116, 22–32. doi: 10.1016/j.bandl.2010.09.012
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* 4, 58–73. doi: 10.1002/(SICI)1097-0193(1996)4:1%3C58::AID-HBM4%3E3.3.CO;2-L
- Xiao, Z., Zhang, J. X., Wang, X., Wu, R., Hu, X., Weng, X., et al. (2005). Differential activity in left inferior frontal gyrus for pseudowords and real words: an event-related fMRI study on auditory lexical decision. *Hum. Brain Mapp.* 25, 212–221. doi: 10.1002/hbm.20105
- Yee, E., Blumstein, S. E., and Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's Aphasia: evidence from eye movements. *J. Cogn. Neurosci.* 20, 592–612. doi: 10.1162/jocn.2008.20056
- Yoncheva, Y. N., Zevin, J. D., Maurer, U., and McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cereb. Cortex* 20, 622–632. doi: 10.1093/cercor/bhp129
- Zemleni, M. Z., Renken, R., Hoeks, J. C. J., Hoogduin, J. M., and Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *Neuroimage* 34, 1270–1279. doi: 10.1016/j.neuroimage.2006.09.048

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 April 2014; accepted: 30 June 2014; published online: 29 July 2014.
Citation: Vitello S, Warren JE, Devlin JT and Rodd JM (2014) Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences. *Front. Hum. Neurosci.* 8:530. doi: 10.3389/fnhum.2014.00530
This article was submitted to the journal *Frontiers in Human Neuroscience*.
Copyright © 2014 Vitello, Warren, Devlin and Rodd. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Listening effort and accented speech

Kristin J. Van Engen^{1*} and Jonathan E. Peelle²

¹ Department of Psychology, Washington University in St. Louis, St. Louis, MO, USA

² Department of Otolaryngology, Washington University in St. Louis, St. Louis, MO, USA

*Correspondence: kvanengen@wustl.edu

Edited by:

Patti Adank, University College London, UK

Reviewed by:

Caroline Floccia, University of Plymouth, UK

Keywords: listening effort, speech comprehension, accent, speech perception, speech perception in noise

Understanding spoken language requires mapping acoustic input onto stored phonological and lexical representations. Speech tokens, however, are notoriously variable: they fluctuate within speakers, across speakers, and in different acoustic environments. As listeners, we must therefore perceive speech in a manner flexible enough to accommodate acoustic signals that imperfectly match our expectations. When these mismatches are small, comprehension can proceed with minimal effort; when acoustic variations are more substantial, additional cognitive resources are required to process the signal. A schematic model of speech comprehension is shown in **Figure 1**, emphasizing that different degrees of acoustic mismatch will require varying levels of cognitive recruitment. Recent research increasingly supports a critical role for executive processes—such as verbal working memory and cognitive control—in understanding degraded speech (Wingfield et al., 2005; Eckert et al., 2008; Rönnberg et al., 2013). However, to date, the literature has focused on sources of increased acoustic challenge that originate in the listener (hearing loss) or in the listening environment (background noise). Largely unexplored are the cognitive effects of accented speech (i.e., speech produced by a speaker who does not share a native language or dialect with the listener), a ubiquitous source of variability in speech intelligibility. Here we argue that accented speech must also be considered within a framework of listening effort.

LISTENING EFFORT

Recent years have seen an increasing focus on the cognitive effects of acoustic

challenge during speech comprehension (Mattys et al., 2012). A common theme is that when speech is acoustically degraded, it deviates from what listeners are used to (i.e., stored phonological and lexical representations), resulting in a mismatch between expectation and percept (Sohoglu et al., 2012, 2014). As a result, listeners must recruit additional cognitive resources to make sense of degraded speech (Rönnberg et al., 2013). Given that listeners' cognitive resources are limited, at some point the allocation of cognitive resources to resolve acoustic challenge will begin to impinge upon other types of behavior. Indeed, even mild hearing loss has been shown to impact syntactic processing (Wingfield et al., 2006), running memory for speech (McCoy et al., 2005), and subsequent memory for short stories (Piquado et al., 2012). Further support for the connection between acoustic and cognitive processing comes from the fact that behavioral challenges are exacerbated in older adults due to age-related cognitive decline (Wingfield et al., 2005).

If increased executive processing is required to deal with acoustic challenge, the effects should not only be apparent in listeners with hearing loss, but in listeners with good hearing in cases of external auditory interference. Consistent with this view, acoustic distortion reduces the episodic recall of word pairs (Heinrich and Schneider, 2011) or word lists (Rabbitt, 1968; Cousins et al., 2014). Conversely, increasing speech clarity through the use of listener-oriented speech facilitates recognition memory for spoken sentences (Van Engen et al., 2012). Thus, listening effort appears to be a general consequence of challenging speech signals, in which

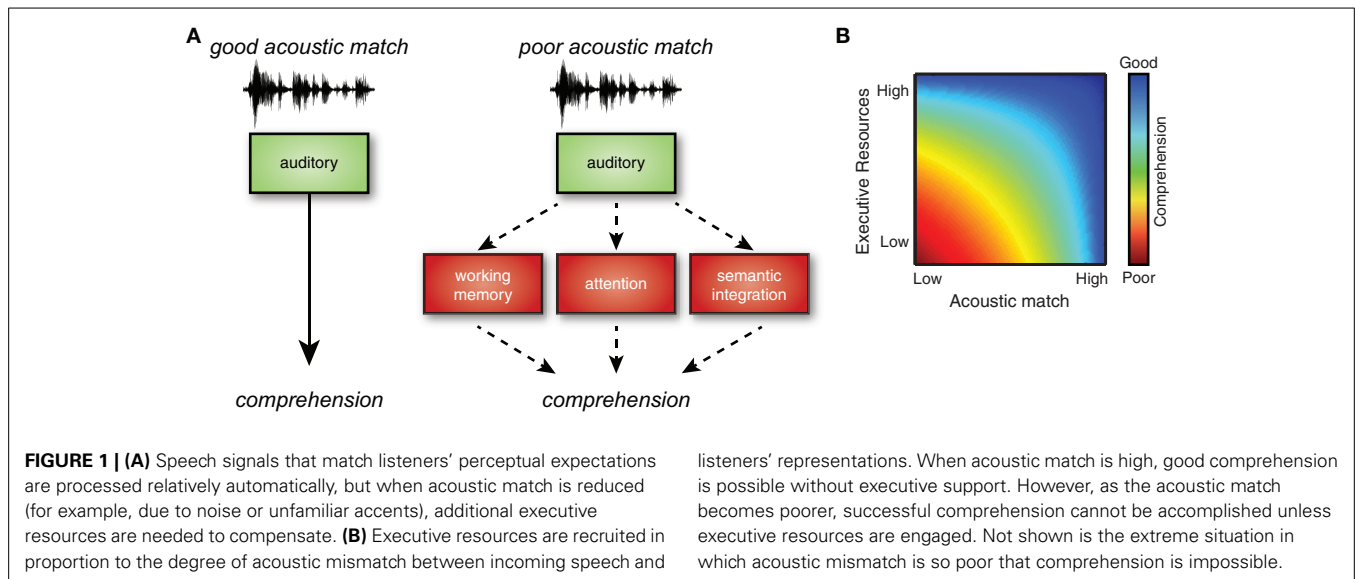
acoustic mismatch can arise due to either internal factors such as hearing loss, or external factors such as background noise.

Functional neuroimaging studies have begun to link these additional executive resources to specific neural systems by identifying increased neural activity resulting from acoustic challenge during speech comprehension (Davis and Johnsrude, 2003; Eckert et al., 2008, 2009; Adank, 2012; Hervais-Adelman et al., 2012; Obleser et al., 2012; Erb et al., 2013). These increases in neural activity frequently involve areas not seen during “normal” speech comprehension—such as frontal operculum, anterior cingulate, and premotor cortex—consistent with listeners' recruitment of additional executive resources to cope with acoustic challenge. Evidence that these increases in brain activity are task-relevant comes from the fact that they vary as a function of attention (Wild et al., 2012), and modulate behavioral performance on subsequent trials (Vaden et al., 2013).

Taken together, then, there is clear evidence that when speech is acoustically degraded, listeners must rely on additional cognitive resources, supported by an extensive network of brain regions. This general principle has been shown in listeners with hearing loss and in good-hearing listeners presented with acoustically degraded materials. In the next section we consider how these findings may play out in the context of understanding accented speech.

LISTENING EFFORT AND ACCENTED SPEECH

If acoustic deviation from stored phonological/lexical representations is indeed the



primary cause of increased listening effort, then speech produced in an unfamiliar accent (whether a regional accent or a foreign accent) should similarly affect not only speech intelligibility, but also the efficiency and accuracy of linguistic processing, and memory for what has been heard. Furthermore, accented speech would also be expected to involve the recruitment of compensatory executive resources.

Foreign-accented speech, for example, is characterized by systematic segmental and/or suprasegmental deviations from native language norms. Naturally, these mismatches can lead to a reduction in the intelligibility of the speech (Gass and Varonis, 1984; Munro and Derwing, 1995; Bent and Bradlow, 2003; Burda et al., 2003; Ferguson et al., 2010; Gordon-Salant et al., 2010a,b). However, even when foreign-accented speech is fully intelligible to listeners (i.e., they can correctly repeat or transcribe it), processing it requires more effort than processing native accents: listeners report that accented speech is more difficult to understand (Munro and Derwing, 1995; Schmid and Yeni-Komshian, 1999), and it is processed more slowly (Munro and Derwing, 1995; Floccia et al., 2009) and comprehended less well than native-accented speech (Anderson-Hsieh and Koehler, 1988; Major et al., 2002). Similar effects have been observed for unfamiliar regional accents: Adank et al. (2009) have shown, for example, that listeners' response times and error rates on

a semantic verification task (i.e., responding to simple true/false questions spoken with different accents) are higher for speech produced in an unfamiliar regional accent. (For a review of the costs associated with processing accented speech across the lifespan, see Cristia et al., 2012.)

The behavioral consequences of listening to accented speech, therefore, include reductions in intelligibility, comprehensibility, and processing speed—all effects that mirror those seen under conditions involving acoustic degradation. To date, there are few functional neuroimaging studies investigating whether increased brain activity is also seen in response to accented speech, although published accounts suggest this is indeed the case (Adank et al., 2012). In general, we would expect that when listening to accented speech, people will recruit comparable executive resources as when listening to other forms of degradation. This would be consistent with increased activity in regions of premotor cortex, inferior frontal gyrus, and the cingulo-opercular network.

That being said, it is important to acknowledge that mismatches between incoming signals and stored representations can arise by different mechanisms. For degraded speech—including steady-state background noise, hearing impairment, or aided listening—listeners experience a loss of acoustic information. This loss is systematic insofar as it involves the inaudibility of a particular

portion of the acoustic signal. In accented speech, there are systematic mismatches between the incoming signal and listeners' expectations, but these arise through phonetic and phonological deviations rather than through signal loss. The degree to which the source of acoustic mismatch affects the type and degree of compensatory cognitive processing required for understanding speech remains an open question. It could be that degraded and accented speech require similar types of executive compensation, and thus both neural and behavioral consequences are largely similar. A second option is that although listeners show similar behavioral consequences to these two types of speech, they are obtained through the use of different underlying neural mechanisms. Finally, there may be differences in both the neural and behavioral consequences of degraded compared to accented speech, or between different types of accented speech. The available preliminary evidence suggests a possible dissociation at the neural level, with different patterns of recruitment for speech in noise compared to accented speech (Adank et al., 2012), and in regional compared to foreign accents (Goslin et al., 2012). However, additional data are needed, and the results may also depend on the level of spoken language processing being tested (Peelle, 2012), task demands, and other factors that determine cognitive challenge for listeners.

ADDITIONAL CONTRIBUTIONS TO LISTENING EFFORT

There are undoubtedly a number of additional influences on the perception of accented speech which may not be relevant for acoustically degraded speech. These include familiarity with an accent (Gass and Varonis, 2006), cultural expectations (Hay and Drager, 2010), and intrinsic listener motivation (Evans and Iverson, 2007). Acoustic familiarity may be specifically related to speech, or simply reflect the experience of a particular listener (Holt, 2006). Together, this confluence of factors can interact with acoustic mismatch to determine the degree of perceptual effort experienced by listeners.

WHY DO THE COGNITIVE CONSEQUENCES OF ACCENTED SPEECH MATTER?

If understanding accented speech indeed requires additional cognitive support, then listeners are likely to have greater difficulty not only understanding their accented interlocutors (i.e., reduced intelligibility), but also comprehending and remembering what they have said, and possibly in managing other information or tasks while listening to accented speech. Given the ubiquity of accented speakers (both foreign and regional) in contemporary society, the practical implications of these problems are wide-ranging. Consider, for example, classrooms with foreign-accented teachers or medical settings where patients and medical personnel who *do* share a language may nevertheless *not* speak with similar accents. In such situations the compensatory cognitive processing that can often (though not always) maintain high intelligibility between speakers and listeners may still come at a cost to listeners' ability to encode critical information. Within the context of a broader framework for effortful listening, it is clear that such challenges will be further exacerbated in the frequently-encountered case of acoustic degradation (such as from background noise or hearing loss), where mismatches between incoming speech and listeners expectations can arise from *both* loss of acoustic information and from distortion due to accent. It has been observed, for example, that noisy or reverberant listening environments disproportionately reduce the intelligibility

of foreign-accented speech as compared to native-accented speech (Van Wijngaarden et al., 2002; Rogers et al., 2006).

An important point is that effortful listening is not an all-or-none phenomenon; rather, the level of cognitive compensation required will depend on the degree of acoustic mismatch in any given listening situation. A relatively mild accent, for example, or one that is highly familiar to a particular listener, can be well understood and require little to no additional effort. Furthermore, we know that listeners can rapidly adapt to both foreign-accented speech (Clarke and Garrett, 2004; Bradlow and Bent, 2008; Sidaras et al., 2009; Baese-Berk et al., 2013) and speech produced in unfamiliar regional accents (Clopper and Bradlow, 2008; Maye et al., 2008; Adank and Janse, 2010). Assuming that understanding accented speech is cognitively challenging due to mismatches between signals and listener expectations, as suggested by the general model of effortful listening presented here, it follows that such perceptual adaptation to an accent will decrease listening effort, and thereby *increase* functional cognitive capacity: Adaptation effectively reduces the mismatch between incoming speech and listener expectations, thus lowering the demand for compensatory executive processes (Figure 1). Auditory training with accented speech may therefore not only be useful for improving intelligibility, but also for increasing listeners' cognitive capacity¹.

CONCLUSIONS

When speech does not conform to listeners' expectations, additional cognitive processes are required to facilitate comprehension. In the case of acoustic degradation, it is increasingly accepted that this type of effortful listening can interfere with subsequent attention, language, and memory processes. Here we have argued that accented speech shares critical characteristics with acoustically degraded speech, and that considering the cognitive consequences of acoustic mismatch is critical

in understanding how listeners deal with accented speech.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the Dana Foundation and the National Institute on Aging of the National Institutes of Health under award number R01AG038490.

REFERENCES

- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: two Activation Likelihood Estimation (ALE) meta-analyses. *Brain Lang.* 122, 42–54. doi: 10.1016/j.bandl.2012.04.014
- Adank, P., Davis, M. H., and Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia* 50, 77–84. doi: 10.1016/j.neuropsychologia.2011.10.024
- Adank, P., Evans, B. G., Stuart-Smith, J. and Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *J. Exp. Psychol. Human.* 35, 520–529. doi: 10.1037/a0013552
- Adank, P., and Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychol. Aging* 25, 736–740. doi: 10.1037/a0020054
- Anderson-Hsieh, J., and Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Lang. Learn.* 38, 561–613. doi: 10.1111/j.1467-1770.1988.tb00167.x
- Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *J. Acoust. Soc. Am.* 133, EL174–EL180. doi: 10.1121/1.4789864
- Bent, T., and Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am.* 114, 1600–1610. doi: 10.1121/1.1603234
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. doi: 10.1016/j.cognition.2007.04.005
- Burda, A. N., Scherz, J. A., Hageman, C. F., and Edwards, H. T. (2003). Age and understanding speakers with Spanish or Taiwanese accents. *Percept. Mot. Skills* 97, 11–20. doi: 10.2466/pms.2003.97.1.11
- Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 101, 2299–2310. doi: 10.1121/1.1815131
- Clopper, C. G., and Bradlow, A. R. (2008). Perception of dialect variation in noise: intelligibility and classification. *Lang. Speech* 51, 175–198. doi: 10.1177/00238309080898539
- Cousins, K. A. Q., Dar, H., Wingfield, A., and Miller, P. (2014). Acoustic masking disrupts time-dependent mechanisms of memory encoding in word-list recall. *Mem. Cogn.* 42, 622–638. doi: 10.3758/s13421-013-0377-7
- Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., and Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Front. Psychol.* 3:479. doi: 10.3389/fpsyg.2012.00479

¹It is possible that improvements over the course of perceptual learning also rely to some degree on executive processes. So although the most straightforward prediction is that perceptual adaption will reduce the demand on executive resources, the degree to which this actually happens is an empirical question.

- Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431.
- Eckert, M. A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., et al. (2009). At the heart of the ventral attention system: the right anterior insula. *Hum. Brain Mapp.* 30, 2530–2541. doi: 10.1002/hbm.20688
- Eckert, M. A., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., and Dubno, J. R. (2008). Age-related effects on word recognition: reliance on cognitive control systems with structural declines in speech-responsive cortex. *JARO* 9, 252–259. doi: 10.1007/s10162-008-0113-3
- Erb, J., Henry, M. J., Eisner, F., and Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *J. Neurosci.* 33, 10688–10697. doi: 10.1523/JNEUROSCI.4596-12.2013
- Evans, B. G., and Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *J. Acoust. Soc. Am.* 121, 3814–3826. doi: 10.1121/1.2722209
- Ferguson, S. H., Jongman, A., Sereno, J. A., and Keum, K. A. (2010). Intelligibility of foreign-accented speech for older adults with and without hearing loss. *J. Am. Acad. Audiol.* 21, 153–162. doi: 10.3766/jaaa.21.3.3
- Floccia, C., Butler, J., Goslin, J., and Ellis, L. (2009). Regional and foreign accent processing in English: can listeners adapt? *J. Psycholinguist. Res.* 38, 379–412. doi: 10.1007/s10936-008-9097-8
- Gass, S., and Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Lang. Learn.* 34, 65–87. doi: 10.1111/j.1467-1770.1984.tb00996.x
- Gass, S., and Varonis, E. M. (2006). The effect of familiarity on the comprehensibility of non-native speech. *Lang. Learn.* 34, 65–87. doi: 10.1111/j.1467-1770.1984.tb00996.x
- Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. J. (2010a). Recognition of accented English in quiet by younger normal-hearing listeners and older listeners with normal-hearing and hearing loss. *J. Acoust. Soc. Am.* 128, 444–455. doi: 10.1121/1.3397409
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Schurman, J. (2010b). Short-term adaptation to accented English by younger and older adults. *J. Acoust. Soc. Am.* 128, EL200–EL204. doi: 10.1121/1.3486199
- Goslin, J., Duffy, H., and Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain Lang.* 122, 92–102. doi: 10.1016/j.bandl.2012.04.017
- Hay, J., and Drager, K. (2010). Stuffed toys and speech perception. *Linguistics* 48, 865–892. doi: 10.1515/ling.2010.027
- Heinrich, A., and Schneider, B. A. (2011). Elucidating the effects of ageing on remembering perceptually distorted word pairs. *Q. J. Exp. Psychol.* 64, 186–205. doi: 10.1080/17470218.2010.492621
- Hervais-Adelman, A., Carlyon, R. P., Johnsrude, I. S., and Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Lang. Cogn. Process.* 27, 1145–1166. doi: 10.1080/01690965.2012.662280
- Holt, L. L. (2006). The mean matters: effects of statistically defined nonspeech spectral distributions on speech categorization. *J. Acoust. Soc. Am.* 120, 2801–2817. doi: 10.1121/1.2354071
- Major, R. C., Fitzmaurice, S. F., Bunta, F., and Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: implications for ESL assessment. *TESOL Q.* 36, 173–190. doi: 10.2307/3588329
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Lang. Cogn. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006
- Maye, J., Aslin, R. N., and Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cogn. Sci.* 32, 543–562. doi: 10.1080/03640210802035357
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R., and Wingfield, A. (2005). Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. *Q. J. Exp. Psychol.* 58, 22–33. doi: 10.1080/02724980443000151
- Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* 32, 12376–12383. doi: 10.1523/JNEUROSCI.4908-11.2012
- Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what “speech” is: a hierarchical perspective. *Front. Hum. Neurosci.* 6:309. doi: 10.3389/fnhum.2012.00309
- Piquado, T., Benichov, J. I., Brownell, H., and Wingfield, A. (2012). The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *Int. J. Audiol.* 51, 576–583. doi: 10.3109/14992027.2012.684403
- Rabbitt, P. M. A. (1968). Channel capacity, intelligibility and immediate memory. *Q. J. Exp. Psychol.* 20, 241–248. doi: 10.1080/14640746808400158
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl. Psycholinguist.* 27, 465–485. doi: 10.1017/S014271640606036X
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031
- Schmid, P. M., and Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *J. Speech Lang. Hear. Res.* 42, 56–64.
- Sidas, S. K., Alexander, J. E. D., and Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *J. Acoust. Soc. Am.* 125, 3306–3316. doi: 10.1121/1.3101452
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32, 8443–8453. doi: 10.1523/JNEUROSCI.5069-11.2012
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 186–199. doi: 10.1037/a0033206
- Vaden, K. I. J., Kuchinsky, S. E., Cude, S. L., Ahlstrom, J. B., Dubno, J. R., and Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *J. Neurosci.* 33, 18979–18986. doi: 10.1523/JNEUROSCI.1417-13.2013
- Van Engen, K. J., Chandrasekaran, B., and Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLoS ONE* 7:e43753. doi: 10.1371/journal.pone.0043753
- Van Wijngaarden, S., Steeneken, H., and Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *J. Acoust. Soc. Am.* 111, 1906–1916. doi: 10.1121/1.1456928
- Wild, C. J., Yusuf, A., Wilson, D., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012
- Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., and Cox, C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *J. Am. Acad. Audiol.* 17, 487–497. doi: 10.3766/jaaa.17.7.4
- Wingfield, A., Tun, P. A., and McCoy, S. L. (2005). Hearing loss in older adulthood: what it is and how it interacts with cognitive performance. *Curr. Dir. Psychol. Sci.* 14, 144–148. doi: 10.1111/j.0963-7214.2005.00356.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 May 2014; paper pending published: 08 June 2014; accepted: 14 July 2014; published online: 05 August 2014.

Citation: Van Engen KJ and Peelle JE (2014) Listening effort and accented speech. *Front. Hum. Neurosci.* 8:577. doi: 10.3389/fnhum.2014.00577

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Van Engen and Peelle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Relationship between perceptual learning in speech and statistical learning in younger and older adults

Thordis M. Neger^{1,2*}, Toni Rietveld¹ and Esther Janse^{1,3}

¹ Centre for Language Studies, Radboud University Nijmegen, Nijmegen, Netherlands

² International Max Planck Research School for Language Sciences, Nijmegen, Netherlands

³ Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands

Edited by:

Carolyn McGettigan, Royal Holloway
University of London, UK

Reviewed by:

Frank Eisner, Radboud University,
Netherlands

Julia Erb, Max Planck Institute for
Human Cognitive and Brain
Sciences, Germany

*Correspondence:

Thordis M. Neger, Centre for
Language Studies, Radboud
University Nijmegen, PO Box 310,
6500 AH Nijmegen, Netherlands
e-mail: thordis.neger@mpi.nl

Within a few sentences, listeners learn to understand severely degraded speech such as noise-vocoded speech. However, individuals vary in the amount of such perceptual learning and it is unclear what underlies these differences. The present study investigates whether perceptual learning in speech relates to statistical learning, as sensitivity to probabilistic information may aid identification of relevant cues in novel speech input. If statistical learning and perceptual learning (partly) draw on the same general mechanisms, then statistical learning in a non-auditory modality using non-linguistic sequences should predict adaptation to degraded speech. In the present study, 73 older adults (aged over 60 years) and 60 younger adults (aged between 18 and 30 years) performed a visual artificial grammar learning task and were presented with 60 meaningful noise-vocoded sentences in an auditory recall task. Within age groups, sentence recognition performance over exposure was analyzed as a function of statistical learning performance, and other variables that may predict learning (i.e., hearing, vocabulary, attention switching control, working memory, and processing speed). Younger and older adults showed similar amounts of perceptual learning, but only younger adults showed significant statistical learning. In older adults, improvement in understanding noise-vocoded speech was constrained by age. In younger adults, amount of adaptation was associated with lexical knowledge and with statistical learning ability. Thus, individual differences in general cognitive abilities explain listeners' variability in adapting to noise-vocoded speech. Results suggest that perceptual and statistical learning share mechanisms of implicit regularity detection, but that the ability to detect statistical regularities is impaired in older adults if visual sequences are presented quickly.

Keywords: perceptual learning, statistical learning, individual differences, aging, working memory, attention switching control, processing speed, vocabulary

INTRODUCTION

Listeners' ability to rapidly learn to understand unfamiliar speech conditions such as accented, disordered or noise-vocoded speech is impressive. Within a few sentences, listeners learn to map a new type of speech input onto their old percept, some improving their speech recognition performance by more than 60% (Eisner et al., 2010). However, listeners show great variability in the amount of such perceptual learning (Eisner et al., 2010). This raises the question which mechanisms underlie perceptual learning.

Perceptual learning can be defined as "relatively long-lasting changes to an organism's perceptual system that improve its ability to respond to its environment" (Goldstone, 1998, p. 585). As listeners are not able to describe the changes that led to their improved perception, perceptual learning is assumed to be a type of implicit learning (Fahle, 2006). A conceptual framework that accounts for changes in the perceptual system is the Reverse Hierarchy Theory (RHT) (Ahissar and Hochstein, 2004). The RHT argues that perceptual learning is a top-down guided process. When a listener is exposed to a novel speech condition,

initial performance fails as the speech input can no longer be readily matched to higher-level representations such as word representations. According to the RHT, prolonged exposure modifies these higher-level representations, which subsequently enables top-down guidance to retune weights at lower levels of the processing hierarchy: the weights of task-relevant input are increased and the weights of task-irrelevant input are pruned. This process of weight retuning starts at the highest level of the hierarchy and continues gradually to the lower levels (i.e., the reverse hierarchy). When lower-level representations have been modified, performance under difficult conditions can be based on accessing these low-level representations. This is illustrated by findings that adaptation to noise-vocoded speech generalizes to novel words (Hervais-Adelman et al., 2008), to non-words (Loebach et al., 2008) and to the recognition of environmental sounds (Loebach et al., 2009). These generalization findings suggest that perceptual learning in speech modifies representations at lower levels of the hierarchy, that is, representations at a sublexical level (Hervais-Adelman et al., 2008; Banai and Amitay, 2012).

The RHT has been influential in explaining behavioral observations in visual and auditory perceptual learning (Nahum et al., 2010; Banai and Amitay, 2012; Cohen et al., 2013; Sabin et al., 2013). However, the RHT does not specify which processes take place in the initial stages of adaptation that enable the perceptual system to identify task-relevant cues in the input and to modify high-level representations. One of the basic principles in the RHT and other models of perceptual learning is the retuning of weights based on the relevance of features or dimensions for the specific task (Goldstone, 1998; Doshier and Lu, 1999; Ahissar and Hochstein, 2004; Petrov et al., 2005). This principle implies that stimuli have to share certain features, which can thus be considered task-relevant, for perceptual learning and for transfer of learning to occur. Accordingly, several studies have highlighted the importance of structural regularities (Cohen et al., 2013) and of stimulus consistencies for perceptual learning (e.g., Nahum et al., 2010). In other words, for learning to occur, participants need to detect specific regularities in the input. Therefore, individual differences in sensitivity to such regularities may indicate why listeners differ in adapting to unfamiliar speech input.

An implicit learning mechanism that has been linked to pattern sensitivity is statistical learning. Statistical or probabilistic learning describes the ability to implicitly extract regularities from an input by detecting the probabilities with which properties co-occur (Misyak and Christiansen, 2012). Statistical learning has gained increasing attention over the past years in language research, as language itself is probabilistic in nature (Auer and Luce, 2005). Accordingly, co-occurrence probabilities of units have been shown to facilitate processing at various linguistic levels (e.g., effects of phonotactic probability; Vitevitch et al., 2004) or transitional probability (e.g., Thompson and Newport, 2007). Statistical learning has been found to be of major importance in language acquisition (Saffran, 2003). Also in adulthood, individual differences in statistical learning have been shown to predict sentence processing performance (Misyak and Christiansen, 2012). Moreover, deficits in statistical learning ability have been reported for various language-related disorders such as specific language impairment (Evans et al., 2009), agrammatic aphasia (Christiansen et al., 2010), and language-based learning disabilities (Grunow et al., 2006). As statistical probabilities are provided and continuously updated by the input, relying on statistical probabilities actually enables language users to adapt to their environment, which is the essential characteristic of perceptual learning. Therefore, the present study aims to investigate whether statistical learning relates to perceptual learning in speech perception. If adaptation to a novel speech condition and statistical learning share general mechanisms of implicit regularity detection, then statistical learning performance in a non-auditory modality using non-linguistic stimuli should predict individuals' perceptual learning for speech comprehension.

Perceptual learning in speech and statistical learning may also draw (partly) on the same underlying cognitive abilities, such as working memory and attention. Therefore, we investigated whether both types of learning could be predicted from general cognitive and linguistic abilities. Ahissar and Hochstein (2004) proposed that attentional mechanisms may be engaged in choosing which neuronal populations pass on task-relevant

information to the higher levels and in increasing the functional weights of these populations. Several frameworks of perceptual learning incorporate the idea that attentional mechanisms are involved in perceptual learning (e.g., Goldstone, 1998; Fahle, 2006; Doshier et al., 2010). A study on frequency discrimination found that perceptual learning even occurred after training with non-discriminable stimuli (Amitay et al., 2006). Apparently, training directed the participants' attentional focus to the relevant stimulus dimension, which was sufficient to access the relevant low-level representations during the test phase (Amitay et al., 2006). Moreover, performance on a selective attention task predicted the amount of learning in adaptation to accented speech (Janse and Adank, 2012). Further evidence that attention is involved in perceptual learning comes from studies in which listeners were simultaneously exposed to noise-vocoded speech and both auditory and visual distractors (Huyck and Johnsrude, 2012; Wild et al., 2012). Only listeners who attended the noise-vocoded stimuli showed improved performance in recognizing noise-vocoded speech. Similar effects of attentional focus arise in tasks of visual statistical learning. When observers are asked to attend to symbols of a certain color in a two-color symbol stream, statistical learning effects unfold for regularities within the attended color but not for regularities within the unattended color (Turk-Browne et al., 2005). These findings imply that only attended features are effectively learned. It has been proposed that training procedures that facilitate participants to *switch* their attention to appropriate perceptual features (e.g., fixed temporal presentation of multiple stimuli, repeated presentation) may particularly enhance perceptual learning (Zhang et al., 2008). Therefore, attention switching control may be involved in the process of distinguishing relevant from non-relevant features in tasks of implicit learning.

Another cognitive ability that may be involved in tasks of implicit learning is working memory, which is required to simultaneously store and process auditory or visual information (Gathercole, 1999). Performance on working memory tasks has been shown to predict performance in various speech reception tasks (for a review see Akeroyd, 2008) and, more specifically, there are indications that working memory relates to perceptual learning performance. Teenaged students with learning and reading disabilities who participated in perceptual learning tasks of frequency and duration discrimination showed improved working memory skills after training (Banai and Ahissar, 2009). Furthermore, the two students who failed to show perceptual learning were characterized by the poorest working memory capacity in the sample. During training, students were repeatedly presented with the same stimuli, which allowed them to access low-level representations, thereby improving frequency and duration discrimination. Thus, working memory may have aided perceptual learning by keeping stimuli accessible (also see Goldstone, 1998). In contrast to these findings, Erb et al. (2012) did not find an association between working memory and individual differences in adaptation to noise-vocoded speech. Note, however, in this study, working memory was measured by tasks that relied on immediate recall and, hence, on short term memory (i.e., non-word repetition task, digit span forward task). Possibly, more complex span tasks, that measure the ability to simultaneously

store and process information, rather than just recall capacity, may be particularly associated with tasks of perceptual learning. With respect to statistical learning, recent studies reported correlations between working memory capacity and performance on implicit sequence learning tasks (Bo et al., 2011, 2012). However, findings regarding the link between working memory and implicit learning of sequences are controversial (for a review see Janacek and Nemeth, 2013) and it has been argued that working memory as an executive resource is not involved in tasks of implicit learning (Kaufman et al., 2010).

An additional cognitive ability that should be considered is processing speed. Processing speed reflects the efficiency of a processing system to perform simple operations (Kaufman et al., 2010) and as a general index of processing efficiency, may be assumed to facilitate perceptual learning. Previous research showed that processing speed correlates with performance on tasks of implicit sequence learning (Salthouse et al., 1999; Kaufman et al., 2010). Higher efficiency of the processing system may be beneficial at various stages of the adaptation process. In the framework of the RHT, processing speed may reduce listeners' time to retrieve high-level representations and to initiate modification processes. Furthermore, processing speed may accelerate the process of weight retuning, thereby gaining faster access to low-level representations.

As the current study focuses on adaptation for spoken language understanding, perceptual learning may also draw on linguistic knowledge. Davis et al. (2005) presented data on how the so-called pop-out effect accelerates the process of perceptual learning: if listeners knew the content of what was going to be said before they actually heard the sentence in its degraded form, this benefited their perceptual learning. In line with the Eureka effect in the RHT, in which a cue regarding the content of the stimulus can trigger direct perception of the stimulus and facilitates strong and long-lasting learning effects (Ahissar and Hochstein, 2004), this pop-out finding suggests that lexical knowledge facilitates access to higher-level representations, thereby initiating top-down processes that aid sublexical retuning (Davis et al., 2005). Accordingly, vocabulary, as a measure of lexical knowledge, has been found to predict the amount of perceptual learning in listeners who were adapting to an unfamiliar foreign-sounding accent (Janse and Adank, 2012), accents being linguistic degradations of the stimulus. If we assume that lexical knowledge aids perceptual learning by guiding the top-down search, effects of lexical knowledge should also arise in non-linguistic speech degradations. Therefore, we investigate whether linguistic knowledge, as indexed by vocabulary knowledge, may also facilitate shifting of attention to relevant features of acoustically degraded speech.

As we want to investigate which cognitive processes are involved in perceptual learning in speech, we also aim to test whether our findings generalize to a heterogeneous group of listeners. Older adults typically form a highly heterogeneous group, as perceptual and cognitive processing undergo changes over the life span. Age-related changes in hearing acuity (Lin et al., 2011), processing speed, capacity on working memory tests, attentional control (for a review see Park and Reuter-Lorenz, 2009) but also lexical knowledge (Ramscar et al., 2014) may therefore help to identify relevant cognitive processes. Importantly, the ability to

adapt to unfamiliar speech input is preserved throughout the life span (Peelle and Wingfield, 2005; Golomb et al., 2007; Adank and Janse, 2010; Gordon-Salant et al., 2010). However, differences in the amount and pattern of perceptual learning over exposure between younger and older adults also indicate changes in the underlying processes. While younger and older listeners show the same amount of learning in the initial adaptation phase, older listeners' performance plateaus earlier in adapting to unfamiliar speech (Peelle and Wingfield, 2005; Adank and Janse, 2010), older adults show less transfer of learning to similar conditions (Peelle and Wingfield, 2005), and exhibit slower consolidation of learning (Sabin et al., 2013). Such differences illustrate that the interdependency between cognitive functions and implicit learning processes may change as a function of age. Cognitive abilities associated with adaptation to unfamiliar speech in younger adults may not be the same as in older adults. In order to gain more insights into individual abilities associated with adaptation to unfamiliar speech across the life span, we tested both younger and older adults.

In sum, this study investigates perceptual learning for spoken language understanding in younger and older adults. We use noise-vocoded speech, an acoustic degradation of the speech signal which simulates the auditory signal of a cochlear implant. In contrast to naturally occurring variability in speech (such as accents), participants do not encounter noise-vocoded speech in everyday life. As a consequence, all participants share the same naïve exposure level. We specifically study whether perceptual learning is associated with a general ability to implicitly detect statistical regularities. By testing participants' probabilistic sequence learning with visual non-linguistic stimuli, we apply a rigorous test of the association between the two types of implicit learning. Additionally, we investigate whether both types of implicit learning are associated with individual differences in attention switching control, working memory, information processing speed or lexical knowledge.

MATERIALS AND METHODS

PARTICIPANTS

In total, 60 younger and 73 older adults participated in the current study. All participants were native speakers of Dutch, neurologically intact and had no history of language disorders. One younger participant was excluded as he showed floor performance throughout the perceptual learning task (i.e., he did not understand the noise-vocoded speech at all). Younger adults were aged between 18 and 29 years (mean age 21 years, *SD* 2.5 years) and older adults were aged between 60 and 84 years (mean age 68.4 years, *SD* 5.7 years). In both age groups, the majority of participants were female (53 out of 59 participants in the younger and 47 out of 73 participants in the older sample). Participants had normal or corrected-to-normal vision. Participants were recruited via the subject database of the Max Planck Institute for Psycholinguistics and were compensated €8 per hour for their time.

AUDITORY, COGNITIVE, AND LINGUISTIC BACKGROUND MEASURES

Auditory measure

Hearing thresholds. Age-related hearing loss is prevalent in older adults (Lin et al., 2011). Poorer hearing may affect perceptual

learning as auditory input contains less detail, thereby interfering with accessing and retuning low-level representations. Participants' auditory function was assessed by measuring air-conduction pure tone thresholds with the aid of an Oscilla USB-300 screening audiometer. As age-related hearing loss particularly affects sensitivity to high frequencies, a high-frequency pure tone average [PTAH] was taken as index of hearing acuity. This PTAH was calculated as the mean hearing threshold over 1, 2, and 4 kHz (instead of the standard PTA over 0.5, 1, and 2 kHz). Only the PTAH of the best ear was entered in the analysis, as all auditory stimuli were presented binaurally. Twenty-seven older participants actually qualified for hearing aids on the basis of their hearing thresholds according to the standard of hearing-aid coverage in the Netherlands (PTAH of the worst ear ≥ 35 dB HL). None of the participants wore hearing aids in daily life, however. Higher thresholds reflected poorer hearing. Mean thresholds at different frequencies per age group are given in **Figure 1**.

Cognitive measures

Working memory. Participants performed a digit span backward task as an index of working memory capacity. The test was a computerized variant of the digit span backward task included in the Wechsler Adult Intelligence Scale Test (Wechsler, 2004) and presented via E-prime 1.2 (Schneider et al., 2002). Participants were asked to report back sequences of digits in reverse order. Digits were presented in a large white font (Arial, font size 100) against a black background. Each digit was presented for 1 s with an interval of 1 s between the consecutive digits of a sequence. Sequence length increased stepwise from two to seven digits and performance on each sequence length was tested on two different trials (all participants were presented with all sequence lengths, regardless of their performance on earlier easier trials). The actual test trials were preceded by two practice trials with a sequence length of three to familiarize participants with the task. Participants had

to recall 12 test sequences in total. Individual performance was operationalized as the proportion of correctly reported sequences (out of 12).

Processing speed. Information processing speed was assessed by means of a digit symbol substitution task. Participants had to convert as many digits as possible into assigned symbols in a fixed amount of time (90 s). The digit symbol substitution task is a paper-and-pencil test that was derived from the Wechsler Adult Intelligence Scale Test (Wechsler, 2004). Performance was measured by the number of correctly converted digits in 90 s, meaning that higher scores reflected higher information processing speed.

Attention switching control. The Trail Making Test was administered to obtain a measure of attention switching control. The paper-and-pencil test contained two parts. In Part A, participants were asked to connect numbers as quickly as possible in ascending order (i.e., 1-2-3...), the numbers being spread randomly over a white page. The Part B page had both numbers and letters randomly spread over the page. Participants now had to alternately join numbers and letters in ascending order (i.e., 1-A-2-B-3-C...). In both parts, 25 items had to be connected and the total time to complete each part was measured. We calculated the ratio between both parts (Part B/Part A) as measure of attention switching control (Arbuthnott and Frank, 2000), thereby taking general slowing into account (Verhaeghen and De Meersman, 1998; Salthouse, 2011). Higher scores indicated higher costs of switching between letters and numbers, therefore, poorer attention switching control.

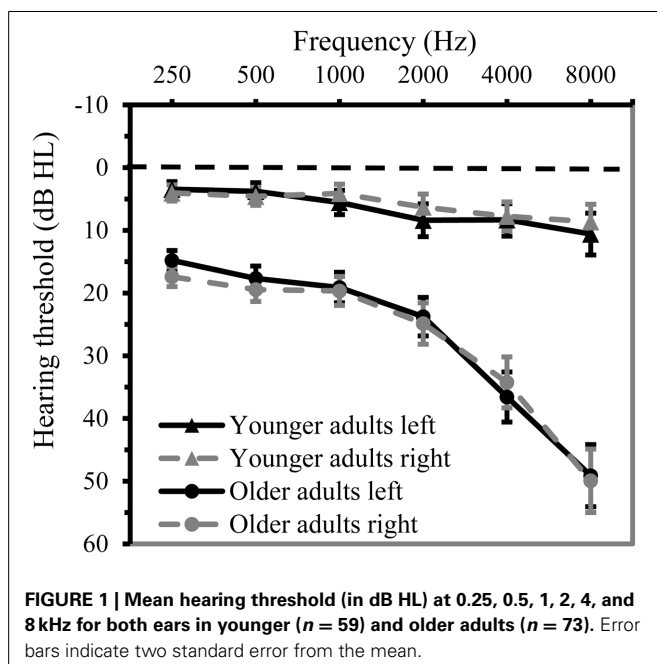
Linguistic measure

Vocabulary knowledge. A vocabulary test in the form of multiple choice questions was administered to obtain a measure of linguistic knowledge (Andringa et al., 2012). The computerized test was administered in Excel (Courier font size 15). Participants had to indicate which out of five possible answers was the correct meaning of Dutch low-frequency words, the last alternative always being "I don't know." Words were not domain-specific and each target word was embedded in a different, neutral carrier phrase. The vocabulary test consisted of 60 items. There was no time limit or pressure to complete the test. Performance was measured by test accuracy, that is, the proportion of correct answers (out of 60). Higher scores thus reflected greater vocabulary knowledge.

STATISTICAL LEARNING

Materials and design

To investigate statistical learning, we adopted the artificial grammar learning—serial reaction time (RT) paradigm (Misyak et al., 2010a). This paradigm has typically been used in studies on statistical learning in language processing and has been found to link to individual language processing abilities (Misyak et al., 2010a,b; Misyak and Christiansen, 2012). As artificial grammar learning simulates language learning processes, the task makes use of auditory presented sound sequences such as non-words. However, as we wanted to investigate whether individuals' ability to adapt to an unfamiliar speech condition could be predicted



by a general ability to implicitly detect regularities, we used visual and non-linguistic stimuli in the statistical learning task. That is, we applied a rigorous test for the relationship between statistical learning and perceptual learning by preventing that a relationship between both measures of learning was specific for auditory and linguistic processing.

Participants were presented with familiar, geometrical shapes in a 2×2 design on the computer screen (see **Figure 2B**), in which one shape on either side of the screen served as target and one as distractor item. Target shapes were sequentially highlighted by a visual marker and participants' task was to click as fast as possible on the highlighted target. The first target was always one on the left side of the screen (i.e., upper or lower one in the first column) and the second target was always on the right side of the screen (i.e., upper or lower one in the second column). The second target was only highlighted after the participant had clicked on the first target item. Crucially, which of the two items in the right-hand column would be highlighted was predictable on the basis of the first target [e.g., in **Figure 2B**, a *triangle* would always be followed by a *star* or a *square* (the latter is not in the display), but never by a *heart*].

Materials consisted of eight familiar, geometrical shapes drawn with a single, continuous black line. The shapes were divided into two grammatical subsets of four shapes each (i.e., Set 1: *triangle*, *hexagon*, *star*, *square*; Set 2: *arrow*, *circle*, *heart*, *cross*). Within each set, two items were selected to appear as first targets (i.e., Set 1: *triangle*, *hexagon*; Set 2: *arrow*, *circle*) and were always followed by one of the other two items that served as second targets (i.e., Set 1: *star*, *square*; Set 2: *heart*, *cross*). Therefore, four combinations of shapes were grammatical within each set, resulting in a total set of eight grammatical combinations (see **Figure 2A**). Target items were presented along with distractors in a rectangular grid display on the computer screen (see **Figure 2B**). Distractor items were shapes from the subset that was currently not tested and the two distractor shapes on the screen formed a grammatical combination themselves. Thus, within a grammatical trial, the transitional probability from the first to the second target was 1, as the first target could only be followed by the target from the same subset. Within the grammar, however, the transitional probability between two adjacent items was 0.5, as a target was followed by a specific successor only half of the time (i.e., a *circle* being followed

by either a *heart* or a *cross*, see **Figure 2A**). Target positions were randomly assigned such that it was unpredictable whether a first or second target would be displayed in the upper or lower row of a particular column.

The artificial grammar learning task was composed of blocks and split into an exposure phase, a test phase and a recovery phase. During the exposure phase, participants could learn the grammar by picking up on the co-occurrence probabilities of the shapes. In total, the exposure phase consisted of 16 grammatical blocks. Within each block, all grammatical combinations were repeated once, resulting in 128 exposure trials (8×16). The test phase consisted of two ungrammatical blocks (2×8 trials). In these ungrammatical blocks, the original grammar was reversed, such that a target was followed by targets of the other (competing) subset. Participants who implicitly learned the grammar should show a drop in performance as they would need to correct their predictions, resulting in a slowed response to the second target. This measure of learning is widely accepted in the literature on implicit learning (Janacsek and Nemeth, 2013): a drop in performance due to removing the underlying regularities can only be linked to grammar sensitivity, whereas learning measures in terms of improvement during the exposure phase cannot be teased apart into general task learning and statistical learning. Therefore, statistical learning was operationalized by the difference in task performance between the last four blocks of the exposure phase (blocks 13–16) and the subsequent ungrammatical test phase (blocks 17–18). The recovery phase again consisted of two grammatical blocks and serves as a control phase. If participants learned the grammar, by re-introducing the regularities in the recovery phase, participants' performance should not decrease any further. In total, the artificial grammar learning task thus contained 20 blocks and 160 trials (8×20).

Procedure

The artificial grammar learning task was presented in E-prime (Schneider et al., 2002) and started with five practice trials that were all grammatical. Participants were instructed to click as quickly as possible on target shapes that were marked by a small filled red cross (10×10 mm) in the center of the target shape. Participants were informed that they had to click on two successive targets and that the first target would be located in

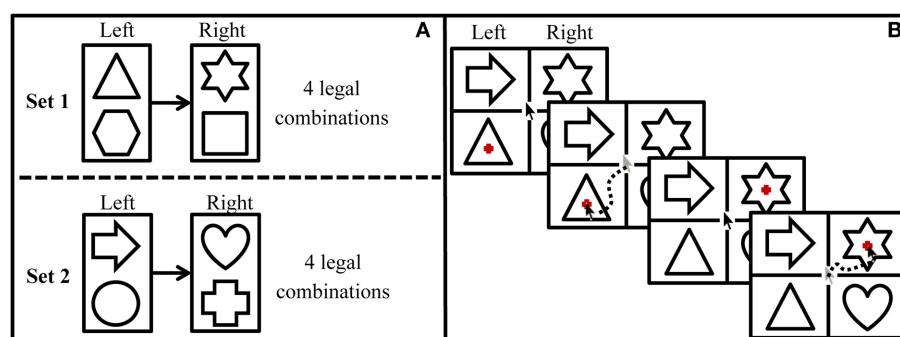


FIGURE 2 | Structure of the statistical learning task. (A) Structure of the grammar in which the first target is always displayed on the left side of the screen and the second target is always displayed on the right side of the screen. **(B)** Procedure of a grammatical trial during the exposure phase.

the first column and the second target would be located in the second column. Each trial started with the presentation of the visual display that consisted of the four shapes and two grid lines, marking the four quadrants on the screen. At the start of each trial, the mouse cursor was located in the center of the screen. Each shape was displayed in a size of 75×70 mm. The visual marker appeared in the middle of the first target shape 500 ms after the onset of the visual display, and was shown until the participant clicked on the marked picture. After the participant had responded, the mouse cursor was automatically set back to the center of the screen to ensure the same distance for all click responses. The second visual marker (same red cross now marking the second target shape) appeared 500 ms after the first click. This time interval was implemented in the design to allow for prediction effects, even in the adults who had slower processing. This time interval had been successfully applied in an earlier study on implicit sequence learning in older adults (Salthouse et al., 1999). Participants could not make errors: the experiment only proceeded if a participant clicked on the appropriate target shape. Clicking on a distractor shape or outside the target picture before giving a correct click resulted in a higher RT. The intertrial-interval was 500 ms. After each block, a small break of 2500 ms was implemented to avoid fatigue effects. During this break, participants saw the block number of the upcoming block and a reminder to click as quickly as possible. It took approximately 20 min to complete the task.

To assess statistical learning, we measured latencies from target highlighting to the subsequent mouse response. Facilitation scores were calculated to index individuals' sensitivity to implicit regularities. The facilitation score was calculated by dividing the RT to the first, unpredictable target within a trial by the RT to the second, predictable target within the same trial. Thus, RT to the first target served as baseline performance within each trial. This was important to minimize biases of task learning and motor performance, particularly for those older adults who may have had little practice in using a computer mouse. During the course of the experiment, RTs may generally get faster as older adults get more experienced in using a mouse. By implementing a new baseline within each new trial, such motor learning should be accounted for. If participants cannot predict which target will be highlighted next, their RTs to both targets within a trial will be similar and will result in a facilitation score of 1. During the exposure phase, learning manifests itself in an increasing facilitation score. That is, if participants learn to predict the second target, RTs to the second item will be faster and, therefore, shorter compared to the first, unpredictable target RTs.

PERCEPTUAL LEARNING

Materials and design

Sixty Dutch sentences were noise-vocoded to create an unfamiliar speech condition to which participants needed to adapt. In noise-vocoded speech, frequency information in the signal is replaced by noise while preserving the original amplitude structure over time. The speech signal was split into multiple non-overlapping frequency bands, which approximately matched equal distances on the basilar membrane (Greenwood, 1990). From each frequency band the smoothed amplitude envelope was derived and

imposed on wide-band noise in the same frequency range. In a last step, these modulated noise bands were recombined, creating a speech signal that sounded like a harsh robot voice. All signal editing was done in Praat (Boersma and Weenink, 2011).

An important characteristic of noise-vocoded speech is that the comprehension level of the speech signal can easily be manipulated by varying the number of frequency bands. The more frequency bands are used to decompose the speech signal, the more detail of the original temporal and amplitude structure is preserved and the more intelligible the speech signal is. Previous research has shown that 10 frequency bands are enough for naïve listeners to immediately understand more than 90% of noise-vocoded speech (Sheldon et al., 2008). However, when presented with speech noise-vocoded with fewer bands, participants only reach this level of performance after a certain amount of exposure.

The maximal amount of learning or intelligibility improvement can be observed if the starting level is neither too high nor too low, so that sufficient information can be derived from the acoustic materials to initiate learning while at the same time allowing for sizeable improvement (see Peelle and Wingfield, 2005). We initially tried to provide participants with an individual starting level from which they could still show improvement. In a separate pilot study, we therefore assigned 23 older adults to a specific noise-vocoding condition (i.e., 4 or 6 bands) on the basis of their performance on a speech reception threshold (SRT) task in noise. Inspection of the data showed that participants' starting level clustered according to band condition. Older adults in the 4 band condition showed a very low starting level (on average they understood only 10% of the sentences correctly), whereas older adults in the 6 band condition showed a very high starting level (on average they already understood 65% of the sentences). Relatedly, the correlation between SRT result and initial performance on the noise-vocoded speech was weak. As our attempt to individualize starting levels on the basis of a speech-in-noise task was not successful, we aimed to provide a roughly similar starting level for both age groups. Based on the results of the pilot study, we decided to present older adults with speech that was vocoded with 5 bands (corner values using 5 frequency bands: 50-280-757-1742-3781-8000 Hz). As younger adults understand more when being exposed to the same degradation as older adults (Peelle and Wingfield, 2005; Sheldon et al., 2008), we presented younger adults with four-band speech (corner values using 4 frequency bands: 50-369-1161-3125-8000 Hz), thus, a more difficult speech condition than older adults (cf. Golomb et al., 2007). Consequently, we were able to see sizeable and comparable amounts of improvement over the course of exposure in both age groups.

Sentences were selected from audiological test materials (Versfeld et al., 2000) and were all produced by the same, male speaker. Each sentence had a length of eight or nine syllables and contained four keywords. Keywords in the selected set of sentences included a noun, verb and preposition. The fourth keyword was an adjective, adverb or a second noun. An example sentence "*De sneeuw glinstert in het maanlicht*" ("*The snow is glistening in the moonlight*") contained the keywords "*sneeuw*," "*glinstert*," "*in*," and "*maanlicht*." Note that five additional sentences were selected for practice purposes, so that there was no overlap

in sentence content between practice and test items. Practice sentences had the same length as test items (a list of all sentences used in the current study is provided in Supplementary Material).

Procedure

An auditory sentence identification task was administered to investigate perceptual learning using the experiment program E-prime (Schneider et al., 2002). Participants listened to the noise-vocoded sentences and were asked to identify and repeat these sentences. They were encouraged to guess if they were unsure. Participants were first presented with five practice trials. First, participants listened to three clear sentences to familiarize them with the task and the speaker. Moreover, these practice trials were used to check whether participants' memory span was sufficient to perform the task given clear input, which was the case for all participants. Then participants listened to two sentences that were noise-vocoded with only two frequency bands to present them with the type of degradation. This more difficult condition with fewer bands was chosen to make sure that no learning could occur during the practice phase (e.g., Ahissar and Hochstein, 1997; Pavlovskaya and Hochstein, 2004; Liu et al., 2008). Practice trials were identical for all participants and were presented in the same order. In contrast, the 60 test sentences were presented in random order for each participant, so that observed learning effects would be independent of inherent intelligibility differences between sentences (e.g., due to differences in semantic predictability). Participants heard a short (125 ms) 3.5 kHz tone to call their attention to the upcoming stimulus 500 ms before sentence onset. After each sentence, the researcher scored the number of correctly repeated keywords (0–4) online. The next trial started immediately after the researcher had confirmed the scoring of the previous trial. Auditory stimuli were presented binaurally via dynamic closed, circumaural headphones (Sennheiser HD 215), at a level of 85 dB SPL. Participants' answers were audiorecorded to allow for later checking of their responses.

EXPERIMENTAL PROCEDURE

Measures of younger adults were obtained in a single experimental session. Testing was spread over two sessions for the older adults, as they also participated in a different study. During the first session, older adults performed the background measures described above. The second session consisted of the statistical learning and the perceptual learning task and followed within a month on the first session. In both age groups, tasks were presented in a fixed order. Although the order differed between younger and older adults, the statistical learning task was always presented before the perceptual learning task. All participants were tested individually in a sound-attenuating booth to minimize distraction. Before the start of each task, participants received verbal and printed task instructions. Participants could ask questions at any time. Between tasks, participants were encouraged to take small breaks.

DATA ANALYSIS

Statistical modeling

To assess learning performance, we implemented linear mixed-effects models using the lmer function from the lme4 package

(Bates et al., 2012) in R (version 2.15.1). In this way, both participants and items could be assessed as random factors and the maximal random slope structure of models could be defined to reduce the probability of a type 1 error (Barr et al., 2013). First, we modeled statistical and perceptual learning performance as a function of age group to assess whether younger and older adults differed in their learning performance. Second, we analyzed the contributions of individual abilities in learning separately within each group as our focus was on individual differences within the respective age groups. Thus, the modeling process that is described here was applied to the statistical learning data and to the perceptual learning data of both age groups.

Linear regression models are based on the assumption that the predictors included in the analysis do not show collinearity (Baayen, 2012). Although some predictor measures were intercorrelated (see Section Performance on Background Measures), we did not control for these intercorrelations for two reasons. First, most correlations explained less than 20% of the variance in the correlated measure (i.e., with correlation coefficients below 0.45). Only the correlation between age and speed in the older adults was moderately correlated ($r = -0.562$). Second, simultaneous inclusion of correlated measures in the analysis has been shown to provide a more reliable interpretation of estimates than inclusion of residualized variables (York, 2012; Wurm and Fissicaro, 2014).

Statistical learning was defined as a drop in performance in the test phase (blocks 17–18) compared to the performance at the end of the exposure phase (blocks 13–16). Therefore, in models of statistical learning, the fixed categorical variable phase (exposure vs. test phase) was the variable of interest to predict individuals' facilitation scores and to indicate learning. Additionally, two control variables and the corresponding two- and three-way interactions with phase were included in models of statistical learning. Control variables were the categorical variable "first target position" (was the first target displayed in the upper or lower row of the left column?) and the categorical variable "target alignment" (were the two targets in a trial aligned horizontally or diagonally?). Given the directionality of Western writing systems, we expected a first target position effect as participants may click faster on a target in the upper left quadrant than in the lower left quadrant. We also expected the drop in facilitation score during the test phase to be less distinct in trials with the first target appearing in the upper left quadrant, such that target position was expected to interact with the amount of learning. Moreover, the alignment of targets was thought to affect second target RTs. Note that the experimental program always set the mouse back to the center of the screen after each click. Despite this automatic mouse reset, participants tended to also move the mouse back to the center of the screen. By doing that, participants unintentionally initiated a movement toward the diagonal shape. Therefore, we assumed that participants would be faster in responding to the second target if targets were arranged diagonally rather than horizontally (see Figure 2), which would result in higher facilitation scores. This direction effect may interact with the effect of removing the regularities, such that the grammaticality effect be decreased for the diagonal movements.

In models of perceptual learning, the number of correctly repeated keywords per sentence served as index of recognition performance and was entered as numerical dependent variable into the model. As perceptual learning was defined as the improvement in speech understanding over exposure, we split the experiment into six blocks, containing 10 sentences each and added block as numerical measure of exposure to the model. However, before block was included in the analysis, we performed a log-transformation of block, as perceptual learning has typically been described by fast initial learning that levels off with increasing exposure (see also **Figure 4**). The transformation of block therefore provided us with an index of exposure that took this non-linear improvement curve into account and converted the improvement over exposure into a linear scale¹.

In the first step of the analysis, we identified the maximal random slope structure of our models to allow for the fact that different participants or items may vary with regard to how sensitive they are with respect to the variables at hand (Cunnings, 2012; Barr et al., 2013): if, e.g., vocabulary knowledge only matters for the understanding of some sentences in the perceptual learning task but not for others, the effect of vocabulary should be modeled individually for each sentence and removed from the fixed effect structure. Changes in the random-slope structure were evaluated by means of the Akaike information criterion (AIC). The model with the lower AIC value (difference ≥ 2) and, therefore, better model fit was retained. As we were interested in the predictors of individual amount of learning, a random participant slope of phase was included in all models of statistical learning. Accordingly, in models of perceptual learning, a random participant slope of block was inserted. That is, all models calculated the learning effect (i.e., the effect of phase in statistical learning and the effect of block in perceptual learning, respectively) individually for each participant.

After determining the maximal random slope structure, we first performed an age group comparison by testing the interactions between age group and the respective index of learning (i.e., phase or block). As younger and older adults may differ with respect to the effects of target position and target alignment on their learning performance, all possible two-way interactions between grammaticality, age group, target alignment and first target position and the three-way interactions between (1) age group, grammaticality and target position and between (2) age group, grammaticality and target alignment were included in the age group comparison of statistical learning.

In a second step, we assessed which cognitive abilities may facilitate learning within the separate age groups. In the statistical learning analysis, the best model that explained the facilitation score on basis of the interactions between phase, target position and target alignment was taken as initial model. In the perceptual learning analysis, the initial model only contained block. Then, measures of age (in older adults only), hearing sensitivity (in

models of perceptual learning only), statistical learning performance (in models of perceptual learning only), attention switching control, working memory, processing speed and vocabulary (all evaluated as numerical covariates) and their interaction with phase (in models of statistical learning) or with block (in models of perceptual learning) were added simultaneously to the initial model. This method of forced entry was preferred, as we had no prior theoretical assumptions about the relative importance of each predictor and aimed to identify those predictors that had unique exploratory power in predicting facilitation scores. All individual predictor measures were centered around their mean prior to inclusion. After we had entered the individual predictor measures, we adopted a backward stepwise selection procedure, in which first interactions and then predictors were removed if they did not attain significance at the 5% level. Each change in the fixed effect structure was evaluated in terms of loss of model fit by means of a likelihood ratio test. Results of the analysis are indicated in estimated absolute effect sizes (β), standard errors, t -values and p -values. Note however that the current version of the lme4 package does not report p -values for t -tests in models with a maximal random slope structure, as it is presently unclear how to calculate the appropriate number of degrees of freedoms (Baayen, 2012). Reported p -values were, therefore, derived by performing a likelihood ratio test between a model that included the specific fixed effect or interaction and a model that did not while all other model parameters were kept constant. That is, p -values actually reflect the significance of loss in model fit if the effect or interaction was excluded from the model.

Individual measure of statistical learning performance

As we wanted to assess whether individual statistical learning performance predicts adaptation to noise-vocoded speech, we needed an index of statistical learning ability for each participant. We derived this index by calculating the random participant slopes of phase (individual adjustments to the general slope) on the basis of the most parsimonious model, in which facilitation scores were predicted only by phase and the control variables (i.e., we derived the measure of statistical learning ability before we included effects of individual predictor measures in the above mentioned analysis).

Thus, we determined an individual value for each participant with which the general effect of phase (in the fixed structure of the model) had to be adjusted to resemble his/her individual learning effect. The lower the value, the more negative was a participant's slope when changing from the end of the exposure phase to the test phase, indicating a steeper drop in facilitation score and, hence, more statistical learning.

RESULTS

PERFORMANCE ON BACKGROUND MEASURES

Mean performance of younger and older adults and age group differences on all background measures are displayed in **Table 1**. As expected, hearing acuity was better in younger adults (i.e., thresholds were lower) than in older adults. Moreover, younger adults showed faster processing and larger memory capacity than older adults. On average, older adults were able to correctly repeat 5.62 test sequences in the working memory test, which

¹Note that we ran a second analysis in which we kept the original index of block. The analysis resulted in the same best models and showed the same effects as the models reported here. However, models that included the log-transformed index of block showed an increased model fit, indicating non-linear learning behavior.

corresponds to a mean digit span of four. Younger adults correctly repeated 8.08 test sequences, corresponding to a mean digit span of five. No difference could be observed in attention switching control between age groups. Older adults outperformed younger adults on the vocabulary test. However, older adults also showed relatively little variation on the vocabulary test [coefficient of variation (SD/M) = 6.9%]. Statistical testing confirmed that the variance in older adults' vocabulary scores was significantly lower than the variability in younger adults' data (coefficient of variation = 11.8%) (Levene's Test: $F = 4.15$, $df_1 = 1$, $df_2 = 130$, $p = 0.044$).

Intercorrelations between background measures within each age group are reported in **Table 2**. In younger adults, significant correlations were observed between the cognitive measures of working memory and processing speed and between working memory and vocabulary. The same intercorrelations were also observed in the older adults. As expected, age correlated with hearing sensitivity and with processing speed within the older sample: older–older participants generally had poorer

hearing and slower processing than younger–older participants. Moreover, processing speed was related to hearing sensitivity in older adults. However, when both measures (i.e., speed and hearing) were controlled for age, this correlation was no longer significant ($r = -0.128$, $p = 0.279$, $df = 71$).

STATISTICAL LEARNING

Valid facilitation scores were restricted to those within 2.5 SD from the mean facilitation score within each age group. **Table 3** shows the average performance of younger and older adults on the statistical learning task in terms of response times and facilitation score. As expected, younger adults were significantly faster in responding to the first target ($t = -84.30$, $df = 23249.45$, $p < 0.001$) and to the second target ($t = -104.34$, $df = 23585.75$, $p < 0.001$) than older adults. Note that all responses in the statistical learning task were accurate as the experimental task only proceeded when a participant had clicked on the correct shape. **Figure 3A** shows the average facilitation scores for both age groups over block². **Figure 3B** displays the mean facilitation scores

Table 1 | Mean performance per age group and age group differences on cognitive, linguistic, and auditory measures.

Measure	Younger adults		Older adults		Age group difference	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Working memory	67.37	17.18	46.80	18.72	6.57	< 0.001
Processing speed	68.10	9.44	48.73	11.01	10.88	< 0.001
Vocabulary	0.68	0.08	0.87	0.06	-15.80	< 0.001
Attention	1.97	0.44	2.05	0.62	-0.96	0.340
Hearing (PTAH)	0.90	5.56	23.31	10.28	-15.96	< 0.001

t-tests tested two-tailed.

²Note that the drop in performance that can be observed in the younger adults during the exposure phase (see **Figure 3A**, blocks 5 and 6) is not significant ($\beta = -0.018$, $SE = 0.010$, $t = -1.71$, $p = 0.088$). This suggests that there was no general drop in performance across the group of younger adults. Likewise, the spread of the individual slope data ($M = -0.018$, $SD = 0.019$, $Min = -0.078$, $Max = 0.027$) also includes positive slope values (indicating

Table 3 | Mean response times (in ms) and facilitation scores of younger adults ($n = 59$) and older adults ($n = 73$) on the statistical learning task.

	1st target RT		2nd target RT		Facilitation score	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Younger adults	499	130	473	129	1.103	0.328
Older adults	705	243	728	240	1.024	0.359

Table 2 | Pearson's correlation coefficients between measures of cognitive, linguistic, and auditory functioning per age group.

Measure	Attention	Memory	Speed	Vocabulary	Hearing
YOUNGER ADULTS ($n = 59$)					
Attention switching	1				
Working memory	-0.109	1			
Processing speed	0.079	0.301*	1		
Vocabulary	-0.071	0.311*	-0.156	1	
Hearing	-0.097	-0.233	0.022	-0.119	1
OLDER ADULTS ($n = 73$)					
Attention switching	1				
Working memory	-0.211	1			
Processing speed	-0.187	0.311**	1		
Vocabulary	-0.076	0.250*	0.207	1	
Hearing	0.104	-0.170	-0.336**	-0.113	1
Age	0.059	-0.194	-0.562**	0.017	0.426**

* $p < 0.05$, ** $p < 0.01$ (tested two-tailed).

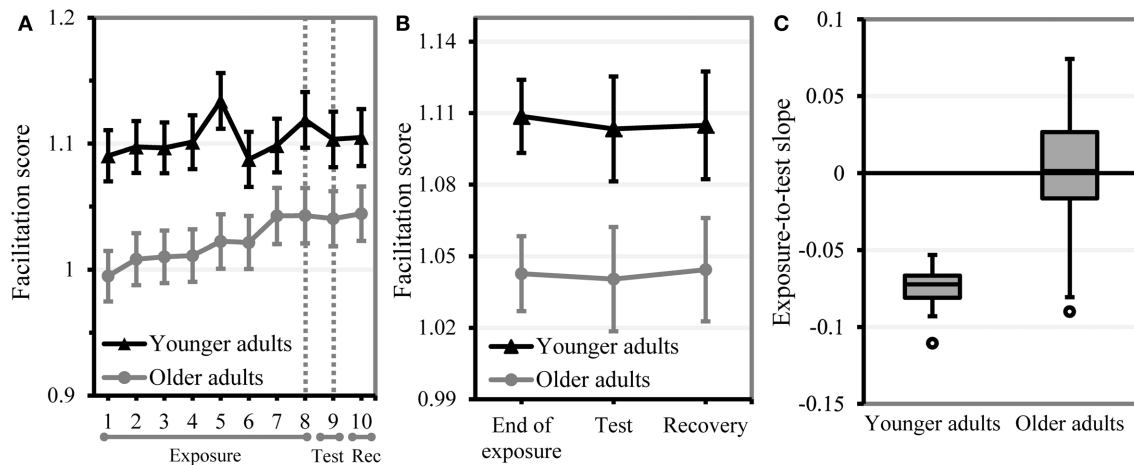


FIGURE 3 | Performance on the statistical learning task. A drop in facilitation score from the end of the exposure phase (blocks 7–8) to the test phase (block 9) indicates learning. Error bars indicate two standard errors from the mean. **(A)** Mean statistical learning performance per age group and block. The area between the dotted lines represents where the effect of

removing the underlying regularities should be observed. **(B)** Mean statistical learning performance per age group and phase. **(C)** Boxplot of statistical learning performance in younger and older adults (individual exposure-to-test slopes from the statistical model). More negative slopes reflect more learning.

at the end of the exposure phase, in the test phase and in the recovery phase to illustrate the learning effect. Moreover, the range of statistical learning that was observed within each age group is displayed in **Figure 3C**. Estimates of the best model within each age group are displayed in **Table 4**.

The age group comparison showed a significant effect of phase ($\beta = -0.137$, $SE = 0.045$, $t = -3.03$, $p = 0.002$), indicating statistical learning in the group of younger adults, who were placed on the intercept. This effect of phase was modified by age group ($\beta = 0.125$, $SE = 0.061$, $t = 2.06$, $p = 0.039$), suggesting that older adults learned less than younger adults and (given the almost equal beta values) that older adults were not affected by removal of the underlying regularities. This interaction between age group and phase tended to be less pronounced in diagonal trials ($\beta = -0.070$, $SE = 0.036$, $t = -1.91$, $p = 0.056$). A fixed effect of age group indicated that, overall, older adults showed a lower facilitation score than younger adults ($\beta = -0.189$, $SE = 0.044$, $t = -4.28$, $p < 0.001$). This effect of age group was influenced by both control variables. That is, the difference in facilitation score between younger and older adults was less distinct in diagonal ($\beta = 0.065$, $SE = 0.025$, $t = 2.54$, $p = 0.011$) and in upper left trials ($\beta = 0.060$, $SE = 0.027$, $t = 2.25$, $p = 0.025$). As expected, facilitation scores were higher in diagonal trials ($\beta = 0.120$, $SE = 0.019$, $t = 6.33$, $p < 0.001$) and lower in trials, in which the first target appeared upper left ($\beta = -0.100$, $SE = 0.020$, $t = -4.98$, $p < 0.001$). Moreover, the effect of phase was modified by both target position and by target alignment, implying that effects of statistical learning were less pronounced in diagonal trials ($\beta = 0.062$, $SE = 0.027$, $t = 2.29$,

$p = 0.022$) and in trials with an upper left target ($\beta = 0.064$, $SE = 0.027$, $t = 2.38$, $p = 0.017$). The random slope structure indicated that participants differed in the degree to which they were affected by target position and by target alignment.

In the younger adults, the best-fitting model showed a significant effect of phase: the facilitation score of younger adults was lower in the test phase than at the end of the exposure phase, indicating that younger adults were affected by removing the underlying regularities. However, none of the individual listener characteristics interacted significantly with test phase, suggesting that amount of statistical learning was not associated with any of the selected measures of cognitive or linguistic abilities. Only processing speed showed a significant fixed effect on facilitation score, indicating that participants with higher processing speed had higher facilitation scores at the end of the exposure phase. As expected, facilitation scores were lower if the first target was displayed upper left and higher if targets were aligned diagonally. Both effects modulated learning in the anticipated direction: the effect of statistical learning was smaller in diagonal trials and in trials in which the first target was displayed upper left. In addition to the random slope of phase, the maximal random slope structure included random effects of first target position and target alignment on participant. Inclusion of these effects suggests that younger participants differed in the degree to which they were affected by target alignment, that is, whether they had to move the cursor horizontally or diagonally. Removing the random slope of phase within subject from the maximal random slope structure did not result in a significant loss in model fit, indicating that the amount of statistical learning did not differ considerably among younger adults (see **Figure 3C**).

Overall, older adults showed no significant effect of test phase, suggesting that they generally did not pick up the subtle regularities in the input. Age was the only individual background measure that predicted performance: the older the participants

improvement, rather than decreased performance). Moreover, a paired samples t-test shows that the size of the unexpected drop in the exposure phase is significantly smaller than the drop in the test phase that is considered to reflect learning ($t = 17.91$, $df = 58$, $p < 0.001$).

Table 4 | Statistical models for the facilitation score of younger and older adults in the statistical learning task.

Fixed effects	Younger adults					Older adults				
	β	SE	t	p		β	SE	t	p	
Intercept	1.098	0.015	73.68	< 0.001	Intercept	1.004	0.015	65.55	< 0.001	
Test phase	-0.069	0.022	-3.09	0.002	Target upper left	-0.102	0.020	-4.98	< 0.001	
Target upper left	-0.099	0.015	-6.78	< 0.001	Diagonal alignment	0.113	0.020	5.77	< 0.001	
Diagonal alignment	0.122	0.020	6.23	< 0.001	Age	-0.004	0.002	-2.36	0.020	
Processing speed	0.002	0.001	2.13	0.034	Target upper left × diagonal	0.133	0.024	5.48	< 0.001	
Test phase × target upper left	0.066	0.025	2.63	0.009						
Test phase × diagonal alignment	0.059	0.025	2.34	0.020						
Random effects										
	Variance	SD	Corr	Corr	Random effects	Variance	SD	Corr	Corr	
Subject	Intercept	0.003	0.055		Subject	Intercept	0.003	0.057		
	Test phase	0.001	0.024	-0.738		Test phase	0.003	0.058	-0.187	
	Diagonal	0.010	0.101	-0.150	0.778	Diagonal	0.007	0.084	-0.300	-0.544
						Lower left	0.003	0.058		
Residual					Residual	Upper left	0.005	0.069	-0.153	
		0.096	0.310				0.120	0.346		

were, the lower was their facilitation score at the end of the exposure phase. In older adults, facilitation score was mainly influenced by the control variables. That is, diagonal alignment of targets enhanced facilitation scores and upper left position of the first target decreased facilitation benefit. A significant interaction between target position and target alignment indicated that effects of one control variable were modified by the other control variable: the effect of the first target being located upper left was smaller when participants could make a diagonal mouse movement to the second target, respectively, the benefit in facilitation score based on a diagonal movement was decreased in case the first target was displayed in the upper left corner of the screen. The maximal random slope structure showed that older adults differed in the degree to which they were affected by changes in target position (random slope of first target position within subject) and target alignment (random slope of first target position within subject). However, in modeling the statistical learning data of the older adults, we had kept in a random slope of phase to allow that participants may vary in how much their performance was affected by removing the regularities (we also needed this random slope parameter as the individual measure of statistical learning). Importantly, inclusion of this random effect of phase did not increase the model fit, implying that older participants did not differ much in their sensitivity to statistical regularities: they were all relatively insensitive to the probabilistic sequence information. Note that older adults continued to show increased facilitation throughout the exposure phase (cf. **Figure 3A**). As their performance was unaffected by the removal of the underlying regularities in the test phase, this suggests that the improvement over block in older adults reflects effects of task learning rather than effects of statistical learning.

PERCEPTUAL LEARNING

As we wanted to include statistical learning performance as a predictor in the analyses of the perceptual learning data alongside the auditory and cognitive measures, we checked for intercorrelations between statistical learning ability and other individual background measures. In the older adults, no correlations were observed. In the younger adults, intercorrelations between statistical learning performance and both working memory ($r = -0.263$, $p = 0.044$; $\rho = -0.297$, $df = 57$, $p = 0.022$) and information processing speed ($r = -0.279$, $p = 0.032$; $\rho = -0.223$, $df = 57$, $p = 0.089$) were significant: more learning was associated with better working memory and with higher processing speed.

In **Figure 4**, the average recognition score per block is displayed to illustrate perceptual learning of the noise-vocoded speech within age group. Moreover, **Figure 4** shows the range of perceptual learning that could be observed within each age group. Although younger adults were presented with a more difficult noise-vocoding condition (4 bands) than older adults (5 bands) and showed a lower starting performance, both age groups showed similar progress in perceptual learning. This indicates that speech conditions were appropriately selected to elicit sizeable and comparable amounts of improvement over the course of exposure in both age groups. Estimates of the best model

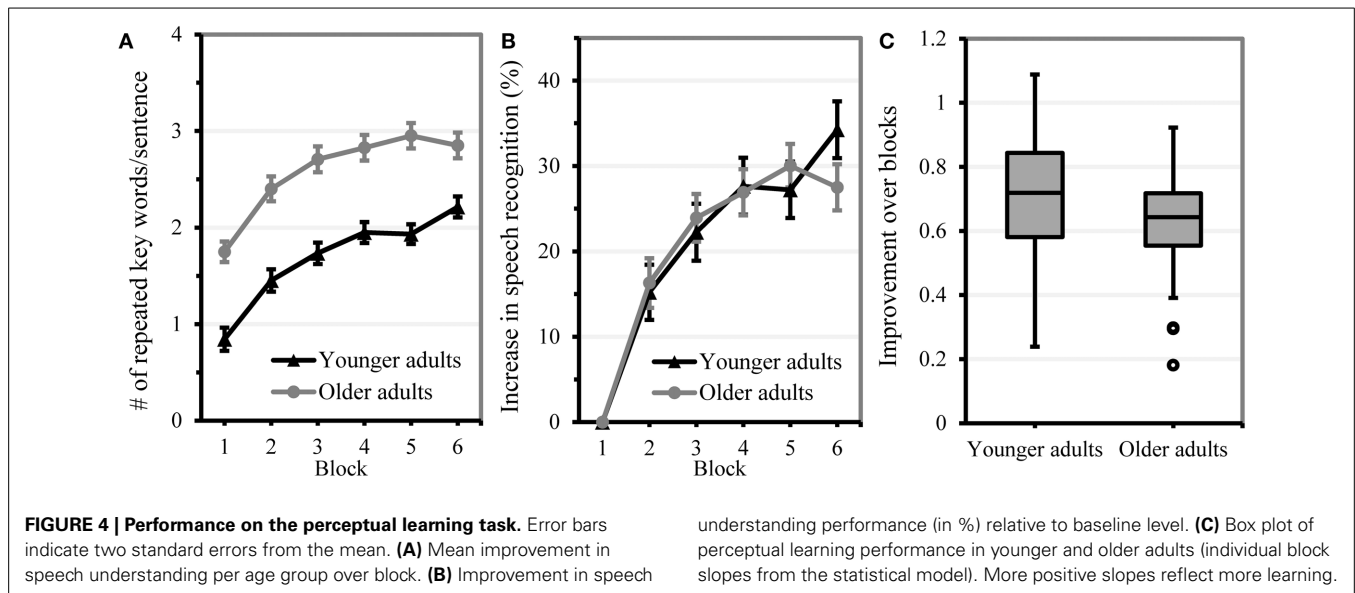


Table 5 | Statistical models for sentence identification performance of younger and older adults in the perceptual learning task.

Fixed effects	Younger adults				Fixed effects	Older adults			
	β	SE	t	p		β	SE	t	p
Intercept	0.917	0.155	5.90	< 0.001	Intercept	1.897	0.122	15.60	< 0.001
Block	0.703	0.043	16.41	< 0.001	Block	0.624	0.037	16.73	< 0.001
Statistical learning	6.649	8.966	0.74	n.s.	Hearing	-0.029	0.005	-6.23	< 0.001
Vocabulary	-0.921	1.255	-0.73	n.s.	Processing speed	0.011	0.004	2.46	0.017
Block \times statistical learning	-9.223	4.058	-2.27	0.023	Age	0.004	0.011	0.38	n.s.
Block \times vocabulary	1.489	0.568	2.62	0.009	Block \times Age	-0.017	0.007	-2.53	0.012
Random effects	Variance		SD	Corr	Random effects	Variance		SD	Corr
Subject	Intercept	0.436	0.660	-0.821	Subject	Intercept	0.122	0.350	-0.509
	Block	0.049	0.221			Block	0.040	0.201	
Item	Intercept	0.912	0.955		Item	Intercept	2.770	1.664	
						Hearing	0.000	0.015	
Residual					Residual	Age	0.001	0.002	-0.887
							1.307	1.143	

n.s. = $p > 0.05$.

to predict sentence identification performance within each age group are displayed in **Table 5**.

The age group comparison showed a significant effect of block ($\beta = 0.710$, $SE = 0.034$, $t = 20.78$, $p < 0.001$) that was not modified by age group ($\beta = -0.071$, $SE = 0.046$, $t = -1.56$, $p = 0.120$), indicating that both age groups showed a similar amount of perceptual learning over the course of the experiment³. As older adults were presented with an easier

condition (5 instead of 4 band vocoded speech), a fixed effect of age group showed that older adults repeated more key words correctly than younger adults ($\beta = 0.971$, $SE = 0.103$, $t = 9.41$, $p < 0.001$). Our results suggest that we were successful in providing older and younger adults with a starting level that

than older adults ($\beta = -0.639$, $SE = 0.160$, $t = 3.99$ and $p < 0.001$). More importantly, in the 4 band condition, we found an interaction between age group and improvement over blocks. That is, older adults showed a significantly smaller effect of block and, hence, less learning, than younger adults ($\beta = -0.098$, $SE = 0.026$, $t = -3.80$ and $p < 0.001$). This result emphasizes that it is not possible to elicit similar amounts of perceptual learning in the two age groups by presenting younger and older adults with the same signal degradation condition.

³Note that we also performed an age group comparison including the data of the pilot study. In this analysis, we compared younger adults' performance on the 4 bands speech (placed on the intercept) to older adults' performance across the different band conditions (i.e., 4 bands, 5 bands and 6 bands). In the 4 band condition, younger adults showed a higher starting performance

allowed for a comparable amount of perceptual learning within both age groups.

In younger adults, none of the predictor measures showed a fixed effect, suggesting that none of the predictor measures could be used to predict initial speech recognition performance (i.e., for block 1 performance, being on the intercept). The best fitting model showed that younger participants identified more keywords correctly with increasing exposure over blocks, indicating that they generally adapted to noise-vocoded speech. This effect of perceptual learning was modified by statistical learning ability: the more participants had picked up the implicit regularities in the statistical learning task (and thus the steeper their drop in performance in the test phase), the more they improved in understanding noise-vocoded speech. This result provides first evidence that perceptual learning and statistical learning are associated. Further, the effect of perceptual learning was modified by vocabulary knowledge: younger adults who had greater vocabulary knowledge showed faster speech adaptation over blocks, underscoring the involvement of linguistic knowledge in perceptual learning of speech. Note that we had excluded an interaction between block and processing speed during the modeling process, as its inclusion led only to a marginal improvement of model fit. This marginal interaction suggested that higher processing speed tended to be associated with faster adaptation. The maximal random slope structure included the effect of block within subject. Removing this effect from the maximal random slope structure reduced the model fit significantly, indicating that individuals differed considerably in perceptual learning ability. As random slopes of individual predictor measures within items did not improve the model fit, this indicated that the effects of predictor measures could be generalized across sentences.

In the older adults, initial sentence identification performance was associated with hearing sensitivity and processing speed: hearing loss considerably affected initial speech understanding, whereas those with higher processing speed showed better initial speech recognition performance. Like the younger adults, older adults showed perceptual learning of noise-vocoded speech, which was indicated by a significant improvement in identification performance over blocks. This block effect was modified by age, indicating that older adults within the older age group improved less over the course of exposure than younger older adults. As age in the older adult sample was intercorrelated with processing speed and hearing sensitivity (see intercorrelations in **Table 2**), we also investigated whether either variable would have surfaced as a predictor for adaptation if we left out age. The variance in amount of perceptual learning that was assigned to age was not taken over by any of the other predictors included in the analysis. This indicates that the effect of age explains unique variance in perceptual learning performance that is not captured by the included cognitive and perceptual predictors. Importantly, statistical learning ability did not facilitate the amount of improvement over the course of the experiment. The maximal random slope structure included effects of age and hearing sensitivity on item, suggesting that the effects of age and hearing sensitivity on recall of noise-vocoded sentences differed across sentences. That is, hearing and age affected speech

understanding of some sentences more than of others, in addition to the general impact these predictors had on sentence recall. Moreover, inclusion of the random effect of block within participant significantly improved the model fit, implying that older participants differed in their improvement to understand noise-vocoded speech over the course of exposure.

DISCUSSION

This study investigated the contribution of general cognitive abilities to listeners' capacity to adapt to novel speech conditions. In order to gain more insight into individual abilities associated with adaptation to unfamiliar speech across the life span, we tested both younger and older adults. Specifically, we aimed to test the hypothesis that listeners' improvement in understanding unfamiliar types of speech could be predicted from individual differences in statistical learning ability and in general cognitive skills.

The ability to implicitly learn has been argued to remain stable over the life span (Midford and Kirsner, 2005). In line with this, several studies reported that older adults are sensitive to probabilistic sequences (Salthouse et al., 1999; Negash et al., 2003; Simon et al., 2011; Campbell et al., 2012) and found the ability to adapt to novel speech conditions to be preserved in older adults (Pelle and Wingfield, 2005; Golomb et al., 2007; Adank and Janse, 2010; Gordon-Salant et al., 2010). Our findings support the notion that perceptual learning ability remains stable over the life span, as both younger and older listeners showed significant improvement in understanding noise-vocoded speech over exposure. Moreover, the observed amount of learning was comparable in both age groups. This suggests that older adults can reach the same amount of perceptual learning as younger adults given better starting level intelligibility. However, only younger adults were sensitive to statistical regularities in the input. As we found a significant learning by age group interaction, this indicated age-related declines in the ability to detect statistical regularities if visual sequences are presented quickly.

Possibly, certain aspects of our statistical learning task may be responsible for the absence of a statistical learning effect in older adults. In particular, we had incorporated an inter-target interval of 500 ms (following Salthouse et al., 1999) between both clicks within a trial to allow for prediction effects, even in older adults with slower processing. As we tested statistical learning in a speeded computer mouse task, and movement control on computer mouse tasks is reduced in older adults (Smith et al., 1999), the implemented inter-target interval may have been too short for older adults to show prediction effects. Moreover, to prevent associations between both measures of implicit learning due to modality-specific processing, we chose for a rigorous test of the association between the two types of learning by testing statistical learning ability in a non-auditory (i.e., visual) domain with non-linguistic stimuli. As older adults were able to implicitly learn in the auditory task, it may be argued that we did not observe implicit learning in the visual paradigm due to age-specific modality effects. In both implicit learning tasks, task-relevant information was presented sequentially (i.e., speech unfolding over time in the auditory task and successive highlighting of targets in the visual task). Visual stimuli have been shown to have

less salient temporal relations than auditory stimuli (Kubovy, 1988). Consequently, auditory learning is superior to visual learning in sequence learning tasks (Conway and Christiansen, 2005). Additionally, a recent study found that statistical learning performance is decreased if visual stimuli are presented at a fast rate (Emberson et al., 2011). Although stimuli presentation in our statistical learning task was not timed as it depended on participants' performance speed (i.e., participants who clicked faster, saw visual stimuli shorter), the time pressure induced by the speeded task, as well as relatively fast and sequential presentation of visual stimuli, may have interfered with statistical learning performance in older adults. That is, results of the current study suggest that older adults' statistical learning ability is affected if fast, sequential processing of visual stimuli is required. However, as previous studies have shown that older adults remain sensitive to probabilistic information in the input (Salthouse et al., 1999; Negash et al., 2003; Simon et al., 2011; Campbell et al., 2012), our failure to observe statistical learning in older adults should not be taken as evidence that older adults are generally insensitive to probabilistic information in the input, or that probabilistic information in the input is generally unimportant for perceptual learning in older adults. Obviously, further research is required to investigate possible links between statistical and perceptual learning in a setting where older adults do show both types of learning.

Overall, limited variability could be observed on the measure of statistical learning ability in both age groups and the amount of individual statistical learning could not be explained by individual differences in cognitive or linguistic abilities in our analyses. However, note the correlations between statistical learning on the one hand and speed and working memory on the other hand in the younger adults. These correlations suggest that, despite relatively little variation in statistical learning, there was some systematicity in younger adults' statistical learning differences. In contrast, participants showed great variability in the amount of adaptation to degraded speech and individual differences in learning to understand noise-vocoded speech could be associated with listeners' cognitive abilities. This finding supports the claim of the RHT that perceptual learning is a top-down guided process, implying that higher cognitive processes are indeed involved in the top-down search to identify task-relevant cues in the input. However, links between cognitive abilities and perceptual learning performance seem to undergo age-related changes, as different associations between perceptual learning ability and cognitive measures emerged in younger and older adults.

In younger adults, initial performance in identifying noise-vocoded speech was not predicted by general cognitive or linguistic abilities. However, differences in the amount of improvement over the course of exposure were associated with individual sensitivity to probabilistic information and with individual vocabulary knowledge. In line with our hypothesis, our results suggest that adaptation to novel speech conditions and statistical learning share mechanisms of implicit regularity detection. Our results contribute to earlier literature indicating a relationship between statistical learning performance and individual differences in language processing (Misyak et al., 2010a). As statistical learning was

tested using visual and non-linguistic stimuli, this suggests that general abilities, that are neither modality-specific nor specific for language processing, drive this association.

As argued in the Introduction, the link between statistical learning and perceptual learning in speech can be twofold. On the one hand, statistical and perceptual learning may be associated as they draw on the same underlying abilities. Our findings do not support this "mediation account": the observed association between perceptual learning in speech and statistical learning performance does not seem to be mediated by the specific cognitive abilities tested in the current study. On the other hand, perceptual learning processes may directly rely on statistical properties in the input. In novel speech conditions, perceptual learning may be facilitated by sensitivity to statistical properties as language itself conveys probabilistic information e.g., in terms of phonotactic (Vitevitch et al., 2004) and transitional probability (Thompson and Newport, 2007). Listeners have been shown to make use of this probabilistic information to segment speech streams into words (Saffran et al., 1997). In the framework of the RHT, listeners who are more sensitive to statistical regularities may, hence, be faster in identifying subunits (e.g., words) in novel speech input, thereby facilitating faster access to high-level representations. Moreover, the information that is transferred from lower to higher levels of the hierarchy may itself be probabilistic in nature. Recent theories in visual perceptual learning argue that the process of input reweighting is based on such probabilistic decisions (e.g., Petrov et al., 2005; Zhang et al., 2010). For example, assuming that the information that is conveyed from lower levels to higher levels is normally distributed and that the mean of the distribution resembles the most relevant input, each incoming input could be reweighted based on its relative distance from the mean, with distance serving as index of informational relevance (Zhang et al., 2010). First evidence that probabilistic information may be encoded in the input from lower to higher hierarchical levels comes from a study in which neuronal network models that relied on probabilistic inferences could explain neurophysiological changes in early sensory areas in visual perceptual learning tasks that could not be accounted for by other models (Bejjanki et al., 2011).

The finding that improvement in understanding noise-vocoded speech in younger adults is predicted by participants' vocabulary size confirms the link between increased lexical knowledge and success in perceptual learning that has previously been reported in adapting to novel-accented speech (Janse and Adank, 2012). Thus, lexical knowledge is not only associated with adaptation to linguistic degradations, e.g., systematic phonological deviations in how a foreign-accented speaker pronounces words, but also relates to perceptual learning of acoustically degraded speech. Previous research has shown that younger and older individuals with higher scores on vocabulary tests also show better performance on measures of verbal fluency (e.g., Kemper and Sumner, 2001; Hedden et al., 2005). Thus, individuals with greater vocabulary knowledge may be more efficient processors of linguistic information (Kemper and Sumner, 2001), and linguistic knowledge may improve perceptual learning in speech by facilitating access to higher-level representations. As access to higher-level representations aids sublexical retuning by enabling

and guiding top-down search processes (Ahissar and Hochstein, 2004), effects of lexical knowledge should in fact arise irrespective of type of systematic speech degradation. Given that Janse and Adank (2012) observed a relationship between vocabulary knowledge and adaptation (to accented speech) in older adults, this raises the question why older adults' perceptual learning performance was not predicted by their linguistic knowledge here. Older adults outperformed younger adults on the measure of lexical knowledge, but note that older adults also showed relatively little variation on the vocabulary test (see Section Performance on Background Measures). Consequently, there was less room to relate lexical knowledge to individual differences in perceptual learning ability in older adults than in younger adults. We checked the variation for older adults' vocabulary scores in the sample of Janse and Adank (2012) (coefficient of variation = 10.3%), which was close to the variation we now observed in the younger adults. Therefore, variation in older adults' vocabulary scores in the current study may have indeed been insufficient to predict perceptual learning.

In the older adult group, listeners' starting level in understanding noise-vocoded speech was associated with higher processing speed and affected by hearing loss, whereas listeners' age predicted how well they adapted to the novel speech condition. That is, younger listeners in the group of older adults showed more learning than older-older listeners. This effect of age had unique explanatory power that was not captured by the included cognitive and perceptual predictors. This finding seems to be consistent with previous research which reported declines in the general identification of noise-vocoded speech with increasing age that were independent of hearing sensitivity (Souza and Boike, 2006; Sheldon et al., 2008) and which may have reflected limited improvement over exposure. Importantly, the current design allowed us to differentiate between effects of individual predictors on both starting level speech identification performance and on amount of perceptual learning. Thus, our results complement earlier findings, suggesting that hearing loss affects initial recognition of noise-vocoded speech, whereas age-related deficits specifically constrain improvement in adaptation to a novel speech input. Younger adults generally outperform older adults when being exposed to the same speech degradation (Pelle and Wingfield, 2005; Sheldon et al., 2008). Importantly, providing younger and older adults with the same speech degradation also has consequences for age groups' ability to improve their performance over exposure (cf. our pilot result data discussed in Section 3.3). In order to have similarly large amounts of perceptual learning for the two age groups, older adults have to be presented with an easier condition than younger adults (Golomb et al., 2007), which was also done in the current study. It is unclear, however, what the age effect on perceptual learning ability among the older adults reflects. A possible account may come from recent studies reporting that coherence between activated brain regions is decreased in older adults (Andrews-Hanna et al., 2007; Pelle et al., 2010), relative to younger adults. Importantly, these deteriorations in connectivity were associated with declines in speech understanding performance under difficult listening conditions (Pelle et al., 2010) and with poorer performance on cognitive tasks (Andrews-Hanna

et al., 2007). In the framework of the RHT, we may speculate that a reduced coordination between neuronal regions may hinder effective information flow between hierarchical levels, thereby constraining processes of input reweighting. Consequently, this decreased information flow would then impede modifications to the lower-level representations. Thus, an age-related decrease in the ability to coordinate activity between brain regions may affect adaptation to challenging novel speech input.

In short, our results suggest that individual differences in general cognitive and linguistic abilities can explain listeners' variability in adaptation to noise-vocoded speech, thereby highlighting the involvement of listener-based abilities in perceptual learning. As noise-vocoded speech simulates the auditory signal of a cochlear implant, findings of the current study may provide valuable insights for aural rehabilitation in younger and older adults. Amount of adaptation over the course of exposure was specifically associated with vocabulary knowledge and with individuals' sensitivity to probabilistic regularities. These combined results emphasize the importance of pattern recognition and linguistic knowledge for perceptual learning and adaptation in speech processing.

AUTHOR CONTRIBUTIONS

Thordis M. Neger was intensively involved in literature search, research design, experiment preparation, data collection, data analyses, data interpretation, and article preparation. Esther Janse was intensively involved in research design, data analyses, data interpretation, and article preparation. Toni Rietveld was involved in data analyses and article preparation, and contributed to the research design. All authors approved the final article.

ACKNOWLEDGMENTS

We wish to thank Willemijn van den Berg for her help in data acquisition. This research was supported by the Netherlands Organization for Research (NWO) under Project No. 276-75-009 (grant awarded to Esther Janse).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2014.00628/abstract>

REFERENCES

- Adank, P., and Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychol. Aging* 25, 736–740. doi: 10.1037/a0020054
- Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature* 387, 401–406. doi: 10.1038/387401a0
- Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464. doi: 10.1016/j.tics.2004.08.011
- Akeroyd, M. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47, 53–71. doi: 10.1080/14992020802301142
- Amiray, S., Irwin, A., and Moore, D. R. (2006). Discrimination learning induced by training with identical stimuli. *Nat. Neurosci.* 9, 1446–1448. doi: 10.1038/nn1787
- Andrews-Hanna, J. R., Snyder, A. Z., Vincent, J. L., Lustig, C., Head, D., Raichle, M. E., et al. (2007). Disruption of large-scale brain systems in advanced aging. *Neuron* 56, 924–935. doi: 10.1016/j.neuron.2007.10.038

- Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., and Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: an individual differences approach. *Lang. Learn.* 62, 49–78. doi: 10.1111/j.1467-9922.2012.00706.x
- Arbuthnott, K., and Frank, J. (2000). Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. *J. Clin. Exp. Neuropsychol.* 22, 518–528. doi: 10.1076/1380-3395(200008)22:4;1-0;FT518
- Auer, E., and Luce, P. A. (2005). “Probabilistic phonotactics in spoken word recognition,” in *The Handbook of Speech Perception*, eds D. Pisoni and R. Remez (Oxford: Blackwell Publishing Ltd), 610–630. doi: 10.1002/9780470757024.ch25
- Baayen, R. H. (2012). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*, 6th Edn. Cambridge: Cambridge University Press.
- Banai, K., and Ahissar, M. (2009). Perceptual learning as a tool for boosting working memory among individuals with reading and learning disabilities. *Learn. Percept.* 1, 115–134. doi: 10.1556/LP.1.2009.1.9
- Banai, K., and Amitay, S. (2012). Stimulus uncertainty in auditory perceptual learning. *Vision Res.* 61, 83–88. doi: 10.1016/j.visres.2012.01.009
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., and Bolker, B. (2012). *lme4: Linear Mixed-effects Models Using Eigen and S4* (Version 0.999-0). Available online at: <http://CRAN.R-project.org/package=lme4>
- Bejjanki, V. R., Beck, J. M., Lu, Z. L., and Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. *Nat. Neurosci.* 14, 642–648. doi: 10.1038/nn.2796
- Bo, J., Jennett, S., and Seidler, R. D. (2011). Working memory capacity correlates with implicit serial reaction time task performance. *Exp. Brain Res.* 214, 73–81. doi: 10.1007/s00221-011-2807-8
- Bo, J., Jennett, S., and Seidler, R. D. (2012). Differential working memory correlates for implicit sequence performance in young and older adults. *Exp. Brain Res.* 221, 467–477. doi: 10.1007/s00221-012-3189-2
- Boersma, P., and Weenink, D. (2011). *Praat: Doing Phonetics by Computer (Version 5.1)*. Available online at: <http://www.praat.org/>
- Campbell, K. L., Zimmerman, S., Healey, M. K., Lee, M. M. S., and Hasher, L. (2012). Age differences in visual statistical learning. *Psychol. Aging* 27, 650–656. doi: 10.1037/a0026780
- Christiansen, M. H., Louise Kelly, M., Shillcock, R. C., and Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition* 116, 382–393. doi: 10.1016/j.cognition.2010.05.015
- Cohen, Y., Daikhin, L., and Ahissar, M. (2013). Perceptual learning is specific to the trained structure of information. *J. Cogn. Neurosci.* 25, 2047–2060. doi: 10.1162/jocn_a_00453
- Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39. doi: 10.1037/0278-7393.31.1.24
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Lang. Res.* 28, 369–382. doi: 10.1177/0267658312443651
- Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J. Exp. Psychol. Gen.* 134, 222–241. doi: 10.1037/0096-3445.134.2.222
- Doshier, B. A., Han, S., and Lu, Z. L. (2010). Perceptual learning and attention: reduction of object attention limitations with practice. *Vision Res.* 50, 402–415. doi: 10.1016/j.visres.2009.09.010
- Doshier, B. A., and Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Res.* 39, 3197–3221. doi: 10.1016/S0042-6989(99)00059-0
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., and Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J. Neurosci.* 30, 7179–7186. doi: 10.1523/JNEUROSCI.4040-09.2010
- Emberson, L. L., Conway, C. M., and Christiansen, M. H. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Q. J. Exp. Psychol.* 64, 1021–1040. doi: 10.1080/17470218.2010.538972
- Erj, J., Henry, M. J., Eisner, F., and Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia* 50, 2154–2164. doi: 10.1016/j.neuropsychologia.2012.05.013
- Evans, J. L., Saffran, J. R., and Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* 52, 321–335. doi: 10.1044/1092-4388(2009/07-0189)
- Fahle, M. (2006). “Perceptual learning,” in *Encyclopedia of Cognitive Science*, ed L. Nadel (Hoboken, NJ: John Wiley & Sons, Ltd.), 1–5.
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends Cogn. Sci.* 3, 410–419. doi: 10.1016/S1364-6613(99)01388-1
- Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612. doi: 10.1146/annurev.psych.49.1.585
- Golomb, J. D., Peelle, J. E., and Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *J. Acoust. Soc. Am.* 121, 1701. doi: 10.1121/1.2436635
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Schurman, J. (2010). Short-term adaptation to accented English by younger and older adults. *J. Acoust. Soc. Am.* 128, 200–204. doi: 10.1121/1.3486199
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species - 29 years later. *J. Acoust. Soc. Am.* 87, 2592–2605. doi: 10.1121/1.399052
- Grunow, H., Spaulding, T. J., Gómez, R. L., and Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *J. Commun. Disord.* 39, 158–170. doi: 10.1016/j.jcomdis.2005.11.004
- Hedden, T., Lautenschlager, G., and Park, D. C. (2005). Contributions of processing ability and knowledge to verbal memory tasks across the adult life-span. *Q. J. Exp. Psychol. Sect. A* 58, 169–190. doi: 10.1080/02724980443000179
- Hervais-Adelman, A., Davis, M. H., Johnsruide, I. S., and Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 460–474. doi: 10.1037/0096-1523.34.2.460
- Huyck, J. J., and Johnsruide, I. S. (2012). Rapid perceptual learning of noise-vocoded speech requires attention. *J. Acoust. Soc. Am.* 131, 236–242. doi: 10.1121/1.3685511
- Janacek, K., and Nemeth, D. (2013). Implicit sequence learning and working memory: correlated or complicated? *Cortex* 49, 2001–2006. doi: 10.1016/j.cortex.2013.02.012
- Janse, E., and Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Q. J. Exp. Psychol.* 65, 1563–1585. doi: 10.1080/17470218.2012.658822
- Kaufman, S. B., De Young, C. G., Gray, J. R., Jimenez, L., Brown, J., and Mackintosh, N. (2010). Implicit learning as an ability. *Cognition* 116, 321–340. doi: 10.1016/j.cognition.2010.05.011
- Kemper, S., and Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychol. Aging* 16, 312–322. doi: 10.1037/0882-7974.16.2.312
- Kubovy, M. (1988). Should we resist the seductiveness of the space-time, vision-audition analogy. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 318–320. doi: 10.1037/0096-1523.14.2.318
- Lin, F. R., Thorpe, R., Gordon-Salant, S., and Ferrucci, L. (2011). Hearing loss prevalence and risk factors among older adults in the United States. *J. Gerontol. A Biol. Sci. Med. Sci.* 66, 582–590. doi: 10.1093/gerona/glr002
- Liu, E. H., Mercado, E., Church, B. A., and Orduna, I. (2008). The easy-to-hard effect in human (*Homo sapiens*) and rat (*Rattus norvegicus*) auditory identification. *J. Comp. Psychol.* 122, 132–145. doi: 10.1037/0735-7036.122.2.132
- Loebach, J. L., Bent, T., and Pisoni, D. B. (2008). Multiple routes to the perceptual learning of speech. *J. Acoust. Soc. Am.* 124, 552–561. doi: 10.1121/1.2931948
- Loebach, J. L., Pisoni, D. B., and Svirsky, M. A. (2009). Transfer of auditory perceptual learning with spectrally reduced speech to speech and non-speech tasks: implications for cochlear implants. *Ear Hear.* 30, 662–674. doi: 10.1097/AUD.0b013e3181b9c92d
- Midford, R., and Kirsner, K. (2005). Implicit and explicit learning in aged and young adults. *Aging Neuropsychol. Cogn.* 12, 359–387. doi: 10.1080/13825580500246894
- Misyak, J. B., Christiansen, M. H., and Tomblin, J. B. (2010a). On-line individual differences in statistical learning predict language processing. *Front. Psychol.* 1:31. doi: 10.3389/fpsyg.2010.00031
- Misyak, J. B., Christiansen, M. H., and Tomblin, J. B. (2010b). Sequential expectations: the role of prediction-based learning in language. *Topics Cogn. Sci.* 2, 138–153. doi: 10.1111/j.1756-8765.2009.01072.x

- Misyak, J., and Christiansen, M. (2012). Statistical learning and language: an individual differences study. *Lang. Learn.* 62, 302–331. doi: 10.1111/j.1467-9922.2010.00626.x
- Nahum, M., Nelken, I., and Ahissar, M. (2010). Stimulus uncertainty and perceptual learning: similar principles govern auditory and visual learning. *Vision Res.* 50, 391–401. doi: 10.1016/j.visres.2009.09.004
- Negash, S., Howard, D. V., Japikse, K. C., and Howard, J. H. (2003). Age-related differences in implicit learning of non-spatial sequential patterns. *Aging Neuropsychol. Cogn.* 10, 108–121. doi: 10.1076/anec.10.2.108.14462
- Park, D. C., and Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annu. Rev. Psychol.* 60, 173–196. doi: 10.1146/annurev.psych.59.103006.093656
- Pavlovskaya, M., and Hochstein, S. (2004). Cross-dimensional transfer and specificity of perceptual learning effects. *Perception* 33, 137–137. doi: 10.1068/v040041
- Peelle, J. E., Troiani, V., Wingfield, A., and Grossman, M. (2010). Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity. *Cereb. Cortex* 20, 773–782. doi: 10.1093/cercor/bhp142
- Peelle, J. E., and Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1315–1330. doi: 10.1037/0096-1523.31.6.1315
- Petrov, A. A., Doshier, B. A., and Lu, Z.-L. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychol. Rev.* 112, 715–743. doi: 10.1037/0033-295X.112.4.715
- Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, H. (2014). The myth of cognitive decline: non-linear dynamics of lifelong learning. *Topics Cogn. Sci.* 6, 5–42. doi: 10.1111/tops.12078
- Sabin, A. T., Clark, C. A., Eddins, D. A., and Wright, B. A. (2013). Different patterns of perceptual learning on spectral modulation detection between older hearing-impaired and younger normal-hearing adults. *J. Assoc. Res. Otolaryngol.* 14, 283–294. doi: 10.1007/s10162-012-0363-y
- Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Curr. Dir. Psychol. Sci.* 12, 110–114. doi: 10.1111/1467-8721.01243
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol. Sci.* 8, 101–105. doi: 10.1111/j.1467-9280.1997.tb00690.x
- Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence* 39, 222–232. doi: 10.1016/j.intell.2011.03.001
- Salthouse, T. A., McGuthry, K. E., and Hambrick, D. Z. (1999). A framework for analyzing and interpreting differential aging patterns: application to three measures of implicit learning. *Aging Neuropsychol. Cogn.* 6, 1–18. doi: 10.1076/anec.6.1.1.789
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools Inc.
- Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (2008). Effect of age, presentation method, and learning on identification of noise-vocoded words. *J. Acoust. Soc. Am.* 123, 476–488. doi: 10.1121/1.2805676
- Simon, J. R., Howard, J. H., and Howard, D. V. (2011). Age differences in implicit learning of probabilistic unstructured sequences. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 66, 32–38. doi: 10.1093/geronb/gbq066
- Smith, M. W., Sharit, J., and Czaja, S. J. (1999). Aging, motor control, and the performance of computer mouse tasks. *Hum. Factors* 41, 389–396. doi: 10.1518/001872099779611102
- Souza, P. E., and Boike, K. T. (2006). Combining temporal-envelope cues across channels: effects of age and hearing loss. *J. Speech Lang. Hear. Res.* 49, 138–149. doi: 10.1044/1092-4388(2006/011)
- Thompson, S. P., and Newport, E. L. (2007). Statistical learning of syntax: the role of transitional probability. *Lang. Learn. Dev.* 3, 1–42. doi: 10.1080/15475440709336999
- Turk-Browne, N. B., Junge, J. A., and Scholl, B. J. (2005). The automaticity of visual statistical learning. *J. Exp. Psychol. Gen.* 134, 552–564. doi: 10.1037/0096-3445.134.4.552
- Verhaeghen, P., and De Meersman, L. (1998). Aging and the Stroop effect: a meta-analysis. *Psychol. Aging* 13, 120–126. doi: 10.1037/0882-7974.13.1.120
- Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J. Acoust. Soc. Am.* 107, 1671–1684. doi: 10.1121/1.428451
- Vitevitch, M. S., Armbruster, J., and Chu, S. Y. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 514–529. doi: 10.1037/0278-7393.30.2.514
- Wechsler, D. (2004). *Wechsler Adult Intelligence Scale. Dutch Version, 3rd Edn*. Amsterdam: Harcourt Test.
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012
- Wurm, L. H., and Fiscaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *J. Mem. Lang.* 72, 37–48. doi: 10.1016/j.jml.2013.12.003
- York, R. (2012). Residualization is not the answer: rethinking how to address multicollinearity. *Soc. Sci. Res.* 41, 1379–1386. doi: 10.1016/j.ssresearch.2012.05.014
- Zhang, J.-Y., Kuai, S.-G., Xiao, L.-Q., Klein, S. A., Levi, D. M., and Yu, C. (2008). Stimulus coding rules for perceptual learning. *PLoS biology*, 6:e197. doi: 10.1371/journal.pbio.0060197
- Zhang, J. Y., Zhang, G. L., Xiao, L. Q., Klein, S. A., Levi, D. M., and Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *J. Neurosci.* 30, 12323–12328. doi: 10.1523/JNEUROSCI.0704-10.2010

Conflict of Interest Statement: The Review Editor Frank Eisner declared to the Guest Associate Editor as being affiliated to the same institution as the authors prior to accepting the review assignment, and the disclosure made was deemed sufficient in this case. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 March 2014; accepted: 28 July 2014; published online: 01 September 2014.
Citation: Neger TM, Rietveld T and Janse E (2014) Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Front. Hum. Neurosci.* 8:628. doi: 10.3389/fnhum.2014.00628

This article was submitted to the journal *Frontiers in Human Neuroscience*.
Copyright © 2014 Neger, Rietveld and Janse. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Effects of navigated TMS on object and action naming

Julio C. Hernandez-Pavon^{1,2,*†}, Niko Mäkelä^{1,2†}, Henri Lehtinen³, Pantelis Lioumis² and Jyrki P. Mäkelä²

¹ Department of Biomedical Engineering and Computational Science, Aalto University School of Science, Espoo, Finland

² BioMag Laboratory, HUS Medical Imaging Center, Helsinki University Central Hospital, Helsinki, Finland

³ Epilepsy Unit, Department of Pediatric Neurology, Helsinki University Central Hospital, Helsinki, Finland

Edited by:

Carolyn McGettigan, Royal Holloway
University of London, UK

Reviewed by:

Joseph T. Devlin, University College
London, UK

Olaf Hauk, MRC Cognition and Brain
Sciences Unit, UK

Patti Adank, University College
London, UK

*Correspondence:

Julio C. Hernandez-Pavon,
Department of Biomedical
Engineering and Computational
Science, Aalto University School of
Science, PO Box 12200, FI-00076
Aalto, Espoo, Finland
e-mail: julio.hpavon@aalto.fi;
julio.hpavon@gmail.com

[†] These authors have contributed
equally to this work.

Transcranial magnetic stimulation (TMS) has been used to induce speech disturbances and to affect speech performance during different naming tasks. Lately, repetitive navigated TMS (nTMS) has been used for non-invasive mapping of cortical speech-related areas. Different naming tasks may give different information that can be useful for presurgical evaluation. We studied the sensitivity of object and action naming tasks to nTMS and compared the distributions of cortical sites where nTMS produced naming errors. Eight healthy subjects named pictures of objects and actions during repetitive nTMS delivered to semi-random left-hemispheric sites. Subject-validated image stacks were obtained in the baseline naming of all pictures before nTMS. Thereafter, nTMS pulse trains were delivered while the subjects were naming the images of objects or actions. The sessions were video-recorded for offline analysis. Naming during nTMS was compared with the baseline performance. The nTMS-induced naming errors were categorized by error type and location. nTMS produced no-response errors, phonological paraphasias, and semantic paraphasias. In seven out of eight subjects, nTMS produced more errors during object than action naming. Both intrasubject and intersubject analysis showed that object naming was significantly more sensitive to nTMS. When the number of errors was compared according to a given area, nTMS to postcentral gyrus induced more errors during object than action naming. Object naming is apparently more easily disrupted by TMS than action naming. Different stimulus types can be useful for locating different aspects of speech functions. This provides new possibilities in both basic and clinical research of cortical speech representations.

Keywords: transcranial magnetic stimulation, speech mapping, left hemisphere, object naming, action naming

INTRODUCTION

Transcranial magnetic stimulation (TMS) is a noninvasive technique where a strong and brief magnetic pulse is delivered to the brain and induces electrical currents. This produces depolarization of cellular membranes and neuronal activation (Barker et al., 1985; Ilmoniemi et al., 1999). TMS has become an important tool for studying speech and language at both the cognitive and neural level (Devlin and Watkins, 2007). TMS may produce both inhibition and facilitation during different phases of speech processing either by directly stimulating a specific speech-related cortical region or indirectly through intracortical networks (Epstein, 1998). TMS has been used for studying the functional localization of speech in healthy subjects, with variable results (Pascual-Leone et al., 1991; Epstein et al., 1999; Devlin and Watkins, 2007; Vigliocco et al., 2011).

Navigated TMS (nTMS) is considered the state-of-the-art technique in performing TMS studies (Siebner et al., 2009). In nTMS, the stimulated cortical site can be defined anatomically from the individual's brain magnetic resonance images (MRI). In addition, orientation and strength of the induced electric field can be estimated (Siebner et al., 2009; Ruohonen and Karhu, 2010). The information provided by nTMS is useful for surgical

planning, and it can be transferred into the operating theater via surgical neuronavigation systems.

So far, nTMS has been used in preoperative localization of the motor cortex (Picht et al., 2009; Vitikainen et al., 2009). It localizes the cortical representations of hand muscles as accurately as direct cortical stimulation (DCS) (Picht et al., 2011; Krieg et al., 2012) and more accurately than functional magnetic resonance imaging (fMRI) (Forster et al., 2011; Krieg et al., 2012). In addition, neuromodulation of Broca's area in speech-related tasks is reported to be more robust by nTMS than by conventional TMS based on external landmarks on the head (Kim et al., 2013). These results motivated us to develop a protocol for preoperative localization of speech-related brain areas by utilizing object naming and nTMS (Lioumis et al., 2012). This novel approach has been compared to DCS during awake craniotomy (Picht et al., 2013). The results imply that nTMS is remarkably sensitive but relatively non-specific in detecting the sites producing speech disturbance in DCS. Discordance between nTMS and DCS was observed particularly in the posterior cortical regions (Picht et al., 2013; Tarapore et al., 2013). Preoperative speech mapping by nTMS can give important a priori information to the neurosurgeons. It may aid in objective preoperative risk-benefit balancing of the planned

surgery, more targeted and smaller craniotomies, faster and safer intraoperative mapping, and safer surgeries for patients that cannot undergo awake craniotomy (Picht et al., 2013). Recently, two studies have used object naming and nTMS to compare language mapping on patients with brain tumors and healthy subjects, suggesting tumor-induced plasticity of speech representation areas (Krieg et al., 2013; Rösler et al., 2013). Thus, better understanding of the effects of TMS during naming tasks may have an impact on surgery planning and provide information about the cortical organization of speech in general.

Picture naming has been extensively studied in both healthy subjects and patients with various neurological diseases. In magnetoencephalography (MEG) and fMRI studies on healthy subjects, action and object naming activate cortical networks including left inferior frontal gyrus, left dorsal premotor, bilateral occipitotemporal, and bilateral parietal areas (Sörös et al., 2003; Petrovich Brennan et al., 2007; Liljeström et al., 2008, 2009). Some functional neuroimaging studies suggest different cortical representations of action and object naming (for a review, see Vigliocco et al., 2011). It has been suggested that action naming activates particularly the left premotor (Valyear et al., 2007; Canessa et al., 2008), parietal (Noppeney et al., 2005), and frontal cortex (Vigliocco et al., 2011), whereas object naming activates the left temporal areas most strongly (Vigliocco et al., 2011). In line, patients with aphasia due to lesions in left frontal areas have shown more severe deficits in action naming, whereas lesions in the left temporal areas are associated with deficits in object naming (Mätzig et al., 2009; Vigliocco et al., 2011).

Action naming appears to be a demanding process that requires more extensive neural processing than object naming (Mätzig et al., 2009). TMS studies indicate that left prefrontal and motor cortices are involved in processing verbs and actions (Pulvermüller et al., 2005; Cappelletti et al., 2008; Gerfo et al., 2008). However, the areas stimulated in (Cappelletti et al., 2008) and (Gerfo et al., 2008) do not match with areas of greater activation for verbs and nouns in imaging studies using similar tasks (Vigliocco et al., 2011). Intraoperative cortical mapping by DCS during awake craniotomy in tumor surgery has systematically revealed widely distributed and highly individual effective cortical sites (Whitaker and Ojemann, 1977; Sanai et al., 2008; Corina et al., 2010) and dissociation of sites inducing errors in action and object naming (Corina et al., 2005; Lubrano et al., 2014). The cortical sites activated specifically by action naming resided mainly in the parietal cortex (Corina et al., 2005).

We mapped the left-hemispheric speech-related areas by nTMS during object and action naming tasks. The induced errors during object and action naming were categorized by type and location of the stimulated cortical site, and compared with each other. We were particularly interested to see if action naming would be interfered more by nTMS in the posterior cortical areas, where discordant results between nTMS and DCS were seen in an object naming task (Picht et al., 2013; Tarapore et al., 2013). If so, action naming tasks might add information in detecting speech-related cortical areas from this region by means of nTMS, as suggested in previous studies (Corina et al., 2005; Noppeney et al., 2005). As action naming is considered more demanding than object naming (Berndt et al., 1997; Mätzig et al., 2009), we

hypothesized that action naming would be more easily disturbed by nTMS than object naming.

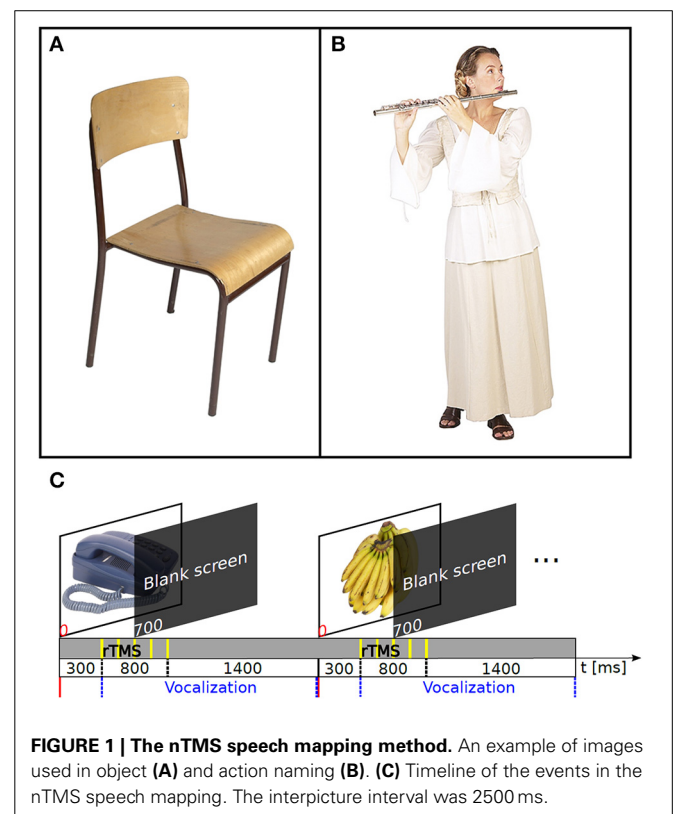
METHODS

SUBJECTS

Eight neurologically normal right-handed subjects (native speakers of Finnish; mean age 26 ± 2 years, four females) participated in the study. The subjects had normal or corrected-to-normal vision. The study was approved by the Ethics Committee of Helsinki University Central Hospital and was in compliance with the declaration of Helsinki. The subjects gave their written informed consent before the experiments.

OBJECT AND ACTION NAMING

We used two sets of color pictures with a white background, one with 131 images depicting objects and another with 98 images depicting actions. Object images illustrated a simple object (e.g., a chair; **Figure 1A**; see also the video in the Supplementary Material). The action images represented a simple event (e.g., playing an instrument; **Figure 1B**). The subjects were asked to name objects or actions in Finnish as quickly and precisely as possible. Two subjects performed action naming before object naming. The experiment consisted of two baseline sessions without nTMS (one for object and another for action naming) and two nTMS sessions (one with object naming and another with action naming). All sessions were video-recorded for offline analysis. The baseline sessions were done before the nTMS sessions. Images that were unfamiliar or named incorrectly in the baseline session were removed from the image set used during nTMS (see



the Supplementary Table). Thus, only fluently named images were used during the nTMS sessions. The numbers of rejected object and action images did not differ significantly (Mann–Whitney *U*-test; $p = 0.26$). The images were displayed in random order within the object naming and action naming sessions (see the **Supplementary Video**). For each subject, all TMS measurements were performed in a row.

STIMULATION

Two recording setups were used. In setup 1, we used eXimia Navigated Brain Stimulation (NBS) version 3.2 (Nexstim Ltd., Helsinki Finland); for details, see Lioumis et al. (2012). In setup 2, we used eXimia NBS version 4.3 and a commercial speech-mapping module (NexSpeech, Nexstim Ltd., Helsinki Finland). Both navigation systems calculate the strength of the maximum electric field that is overlaid on-line on the 3-D reconstruction of the individual's brain (Ruohonen and Karhu, 2010). Each stimulation site is tagged to the MR image for subsequent studies.

All stimulations were done with a biphasic figure-of-eight coil. The outer diameter of the coil was 70 mm. The resting motor threshold (MT) was determined from the right abductor pollicis brevis (APB) muscle, and the strength of the induced electric field at the cortex was registered. These electric fields varied between 40 and 100 V/m at approximately 25 mm from the head surface (i.e., at the navigation depth). The stimulus intensity for the speech mapping was adjusted to produce roughly as strong electric field to perisylvian cortical regions. Navigated TMS of temporal areas occasionally produces some discomfort. However, we were particularly meticulous to avoid such unpleasantness; if the stimulation caused discomfort to the subject due to muscle contraction (in a short test session before the actual measurements), the stimulation intensity was lowered in decrements of 5–10% until it was tolerable. Moreover, the experimental setup has been validated by DCS, where discomfort due to scalp or muscle stimulation is not an issue, and found to match well with DCS particularly in ventral anterior areas (see e.g., Picht et al., 2013; Tarapore et al., 2013; Krieg et al., 2014). Consequently, the stimulation intensity varied somewhat across subjects (80–110% of the APB MT; 30–40% of the stimulator output). The stimulation was done with nTMS trains of five pulses at 5 Hz (Epstein et al., 1996; Lioumis et al., 2012). The subjects wore earplugs during all sessions.

The object and action pictures were displayed for 700 ms on a computer screen once every 2.5 s. The nTMS trains were delivered with a 300 ms delay after the picture onset (**Figure 1C**; **Supplementary Video**). The nTMS onset time was chosen on the basis of MEG studies on dynamics of cortical language processing (Salmelin et al., 2000; Sörös et al., 2003); essentially we did not want to interfere with the visual inspection, but to disturb other stages of language production (e.g., conceptual processing, lexical selection, phonological encoding, and articulatory preparation). The coil was hand-held and it was moved freely between the pulse trains. Approximately 200 sites were stimulated in the left hemisphere by moving the coil semi-randomly in between the trains of pulses, following a grid-like pattern so that the tested target sites covered systematically a wide fronto-temporo-parietal cortical area. The same areas were stimulated for both tasks. The orientation of the coil was adjusted to induce current primarily perpendicular to the fibers of the temporalis muscle to minimize

muscle twitching, and secondarily perpendicular to the sulcus at the stimulation target. The cortical sites where nTMS-induced errors were observed online and were revisited to evaluate the repeatability of the effect (see **Supplementary Video**). On average, 257 stimulus trains were delivered to the left hemisphere during object naming and 243 during action naming in each subject. The maximum difference between repetitions for two different images was one, as the content of the subject-validated image stack was randomized each time a new round of the images started.

DATA ANALYSIS

A neuropsychologist with expertise in effects of DCS on speech (HL) analyzed naming performance in the recorded videos. During the analysis, the stimulation sites were not visible. The baseline naming responses were compared with those recorded during nTMS. The observed errors were categorized as no-response errors, semantic paraphasias, and phonological paraphasias according to previous studies (Corina et al., 2010; Picht et al., 2013; Rösler et al., 2013). *No-response errors*: stimulation leads to a complete lack of naming response. *Phonologic paraphasias*: characterized by unintended phonemic modification of the target word. The spoken word resembles the target word, but is phonetically different. For example the target word “pants” is replaced with “plants.” *Semantic paraphasias*: errors in which the patient substitutes a semantically related or associated word for the target word. For example, the target word “cow” is replaced by the word “horse.” When a naming error occurred, the corresponding nTMS location was marked as speech-related and tagged by the observed error type. Thereafter, the nTMS sites eliciting naming errors were grouped into cortical regions. For the anatomical labeling, we used the anatomical atlas shown in Corina et al. (2010) as in previous publications (Lioumis et al., 2012; Picht et al., 2013). The cortical surface of each subject was separated into anatomical regions according to this template.

The statistical significance of the results were evaluated both in single-subject and group level. For the single-subject analysis, the statistical significance of the observed effects of nTMS on performance in the naming tasks was evaluated separately for each subject and stimulated area. The null hypothesis was that the observed errors occur due to chance. If so, the number of observed errors should follow a Poisson distribution with the parameter $\lambda = \text{number of observed errors (per area)}/\text{the total number of nTMS trials (per area)}$. The probability that the observed number of naming errors in an area could have arisen by chance rather than due to the effect of nTMS was computed by comparing the number of observed errors with one million simulated Poisson samples. The number of samples in the simulated data that were greater than or equal to the observed number of errors gives the probability of the case that the observed effect could have occurred by chance. The significance level was set at 5%. False discovery rate (FDR) was applied on the *p*-values collected from the area wise analyses of each subject to correct for multiple comparisons (Storey, 2002).

For the group level analysis, the 2-tailed Mann–Whitney *U*-test was used to compare the number of errors during object and action naming. The statistical analysis was done to the total

number of naming errors in the left hemisphere, and within each error type and gyrus. The significance level was set at 5%.

To visually summarize the speech mapping results, the stimulation sites that were associated with naming errors from all eight subjects were projected on the standardized MNI brain

template (Mazziotta et al., 2001), using FSL (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012) and FreeSurfer (Fischl et al., 1999) softwares. The brain was segmented from the individual T1-weighted MRIs of each subject and registered with the standard brain template in MNI space. Thereafter, the

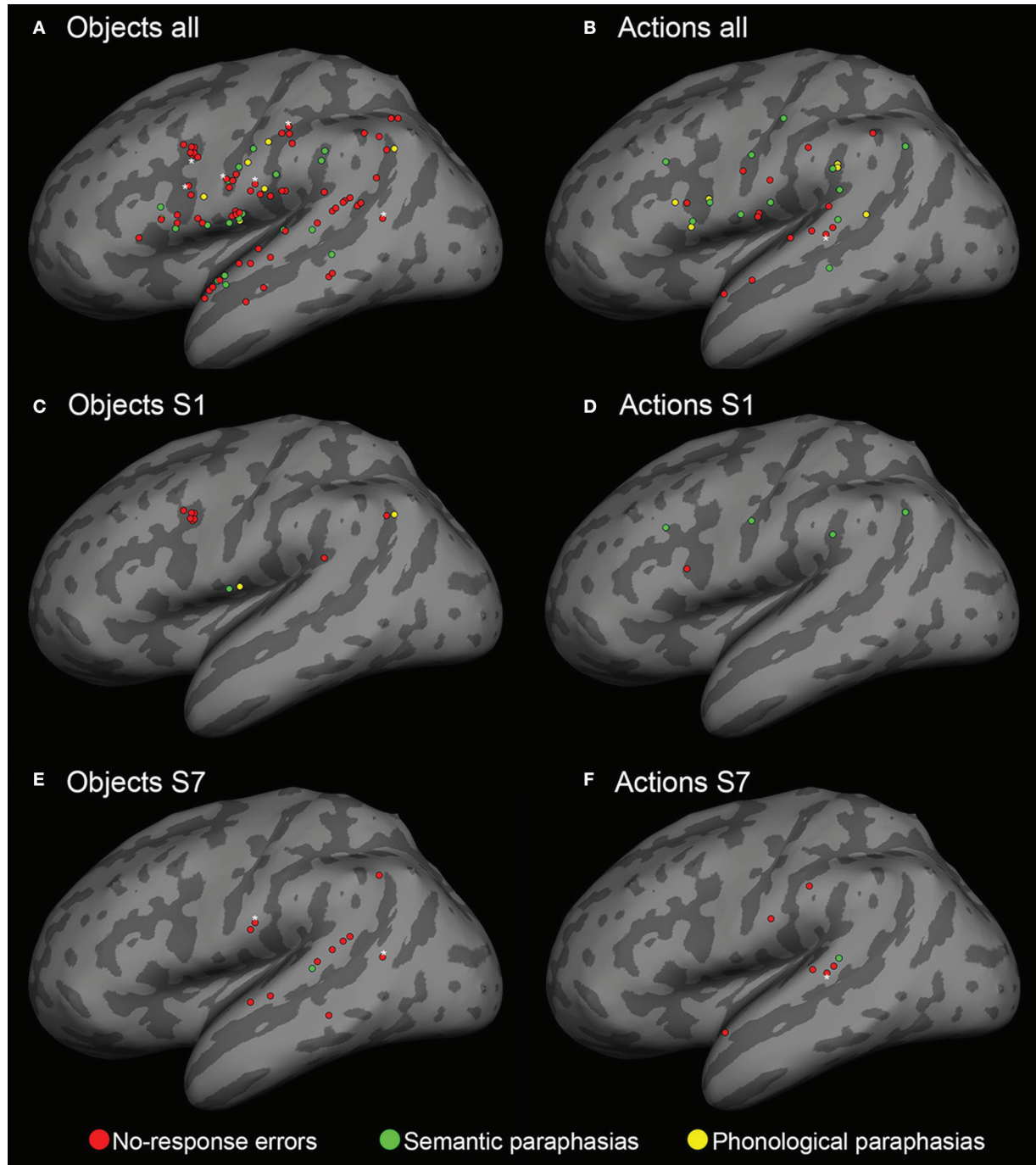


FIGURE 2 | Cortical sites for both object and action naming errors visualized on an inflated reconstruction of the cortex. Red spheres: no-response errors; green spheres: semantic paraphasias; yellow spheres: phonological paraphasias. (A,B) All cortical sites that elicited nTMS-induced

naming errors in the subjects. (C–F) Individual data from subjects S1 and S7. The number, type, and location of the naming errors vary between subjects. The white asterisks indicate the sites of repeated errors at the same location.

Table 1 | Observed nTMS-induced errors for all subjects during object and action naming in different cortical areas (object, action).

Areas/Subjects	No-response errors								Semantic paraphasias								Phonological paraphasias								Total errors	
	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8	S1	S2	S3	S4	S5	S6	S7	S8		
anG	1;0	1;0	0;0	0;0	0;0	0;0	0;0	0;0	0;1	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	2;1	
IFG	2;1	3;0	3;0	0;0	0;0	1;0	0;1	7;0	0;1	0;0	0;0	0;0	0;0	0;2	0;0	1;0	0;0	0;0	0;0	0;0	0;1	1;1	0;0	0;0	18;7	
MFG	2;0	1;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	3;1	
PoG	0;0	2;0	0;0	0;0	2;0	2;0	3;2	5;1	1;0	1;0	0;0	1;0	0;0	1;0	0;0	1;0	1;0	0;1	0;0	0;0	0;0	3;0	0;0	0;0	23;4	
PrG	2;0	1;0	4;1	0;0	0;0	0;0	0;0	1;0	0;1	1;1	0;0	0;1	0;0	1;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	1;0	0;0	0;0	11;4	
STG	0;0	0;0	2;3	6;0	1;1	0;0	5;4	2;0	0;0	0;0	3;1	2;0	0;0	0;0	0;1	0;0	0;0	0;1	0;0	0;0	0;0	0;0	0;0	0;0	21;11	
MTG	0;0	0;0	1;0	0;0	0;0	0;0	5;0	0;0	0;0	0;0	1;1	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	0;0	7;1	
SMG	1;0	1;0	0;1	0;0	0;0	0;0	0;0	1;0	0;1	1;1	0;0	0;0	1;0	1;0	1;0	0;0	1;0	0;1	0;0	0;0	0;0	0;0	0;0	0;0	8;4	
Total errors (object,action)	8;1	9;0	10;5	6;0	3;1	3;0	13;7	16;1	1;4	3;2	4;2	3;1	1;2	3;0	1;1	2;0	2;0	0;3	0;0	0;0	0;1	5;2	0;0	0;0	93;33	
																									7;6	
																									18;12	
																									68;15	

Notation: anG, angular gyrus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; PoG, postcentral gyrus; PrG, precentral gyrus; STG, superior temporal gyrus; MTG, middle temporal gyrus; SMG, supramarginal gyrus.

coordinates of the naming error locations were projected into the MNI space, using the transformation matrix given by the registration and overlaid with the inflated cortical surface of the MNI brain template (Figure 2).

RESULTS

Overall 93 nTMS trains (4.5% from a total of 2056 trains) induced errors during object naming. During action naming, 33 nTMS trains (1.7% from a total of 1944 trains) induced errors (Figure 3A). In seven out of eight subjects, TMS elicited more object naming than action naming errors. In one subject, the total number of induced errors was equal in both (Table 1). Naming errors were induced when nTMS was delivered to angular gyrus (anG), inferior frontal gyrus (IFG), middle frontal gyrus (MFG), postcentral gyrus (PoG), precentral gyrus (PrG), superior temporal gyrus (STG), middle temporal gyrus (MTG), and supramarginal gyrus (SMG) (Table 1 and Figure 2).

In the object naming task, 25% of the sites associated with naming errors were located in the PoG, 23% in the STG, 19% in the IFG, 12% in the PrG, 9% in the SMG, 8% in the MTG, 3% in the MFG, and 2% in the anG (Table 1 and Figure 2). A subanalysis by type showed that 24% of the no-response errors (73% of all naming errors, see Figure 3B) were induced from the IFG, 24% from the STG, 21% from the PoG, and 12% from the PrG (Table 1). Thus, 81% of all sites producing no-response errors during nTMS were concentrated on these areas.

In the action naming task, 34% of the sites associated with naming errors were located in the STG, 21% in the IFG, 12% in the PoG, 12% in the PrG, 12% in the SMG, 3% in the anG, and 3% in the MFG and in the MTG (Table 1 and Figure 2). A subanalysis by type showed that 53% of the sites associated with no-response errors (46% of all naming errors, see Figure 3C), were located in the STG, 20% in the PoG, and 13% in the IFG (Table 1). Thus, more than 80% of all sites producing no-response errors during nTMS were concentrated on these areas. Figure 2 depicts the sites where nTMS elicited naming errors in object and action naming tasks. Fused results for all subjects are shown in Figures 2A,B. Results for two individual subjects are shown in Figures 2C–F to reveal the inter-subject variability. Overall, the number, type, and location of the naming errors varied between the subjects.

The area-dependent subject-level analysis showed significant effects of nTMS in IFG, MFG, PoG, PrG, STG, MTG, and SMG for object naming ($p < 0.05$; see Table 2) and in IFG, PoG, STG, and SMG for action naming ($p < 0.05$; see Table 2). The most sensitive cortical sites were IFG, PoG, PrG, STG, and SMG (see Table 2 for summary). The largest difference of nTMS-sensitive sites in object and action naming tasks was in the PoG, where 7 subjects had a significant effect of nTMS on object naming and only one on action naming (Table 2). Clear individual differences between the subjects in the distributions of the speech-related areas were evident (Table 1 and Figure 2).

In group-level analysis, the total number of nTMS-induced errors in object naming was significantly larger than in action naming ($p = 0.002$; see Table 1). No-response errors were significantly more frequent in object than action naming ($p = 0.002$); the number of semantic and phonological paraphasias did not differ significantly between the tasks. When the total

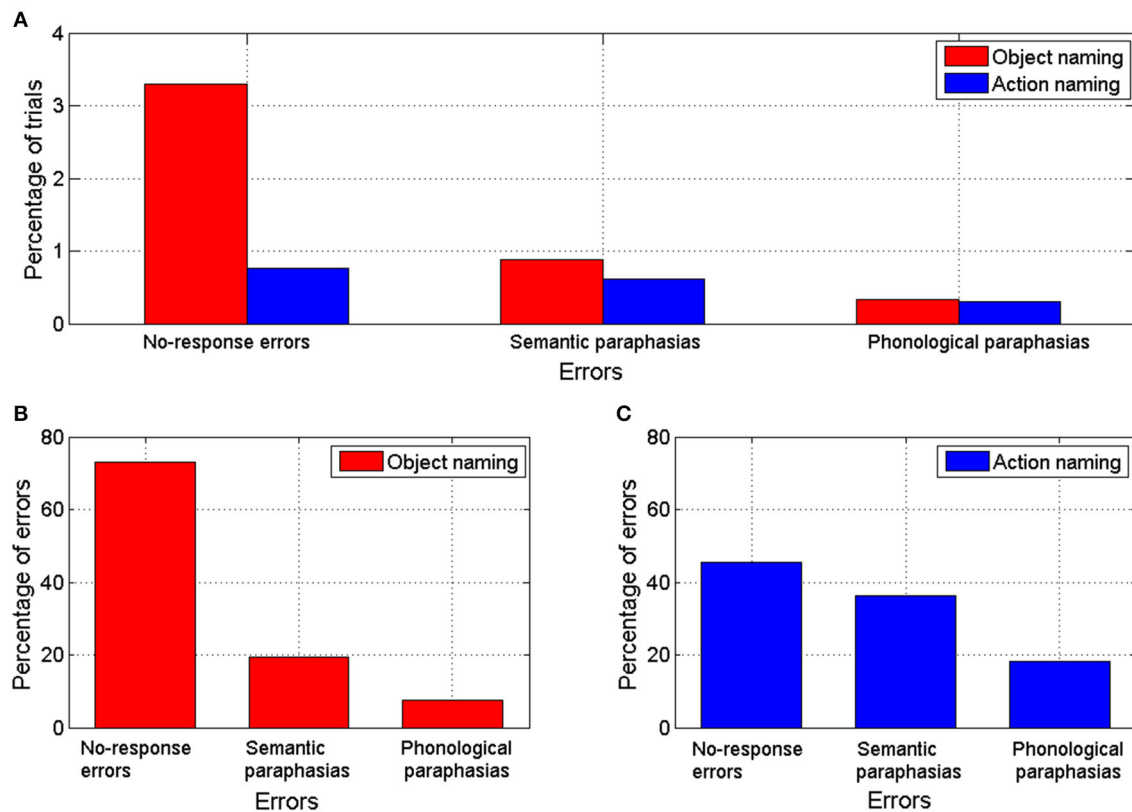


FIGURE 3 | (A) Percentage of trials with errors out of the total number of trials during nTMS for object and action naming tasks. **(B)** Percentage of errors of each error category out of the total number of naming errors

produced by nTMS during object naming task. **(C)** Percentage of errors of each error category out of the total number of naming errors produced by nTMS during action naming task.

number of errors within each gyrus was compared, object naming was more effectively disturbed by nTMS in PoG ($p = 0.014$) than action naming. No significant differences were observed for nTMS in the other gyri (Table 1).

DISCUSSION

We observed that object naming was consistently more disturbed by nTMS to the left hemisphere than action naming. The induced error types varied between subjects, but no-response errors were the most frequent in both tasks. In parallel with our results, object naming errors were more frequent than action naming errors during left-hemisphere DCS of neurological patients (Lubrano et al., 2014). Apparently, object naming is more sensitive to perturbations elicited by nTMS than action naming. DCS is probably more efficient than nTMS; in our study 3.2% of all trials induced naming errors, whereas 11.5% of the tested DCS sites were associated with induced language interferences (Lubrano et al., 2014). However, DCS mapping is limited by the extent of the craniotomy, and our nTMS speech mapping covered a wide cortical area. Hence, it is more likely to stimulate sites that are not speech-related in the nTMS than DCS mapping.

TMS induced naming errors from virtually all perisylvian sites (Figure 2). However, across subjects the location of these individual punctuate regions varied and there were no regionally specific effects of nTMS, which is in line with the results obtained by DCS

studies (Corina et al., 2005; Lubrano et al., 2014). The classical Broca's area (Brodmann area 44/45) in IFG and the Wernicke's area (Brodmann area 22) in STG were both sensitive to nTMS in most subjects. It is evident, however, that the classical modular brain-language model is insufficient to explain our results. Instead, the results support the current state-of-the-art models of widely distributed language network (Poeppel and Hickok, 2004; Hagoort and Indefrey, 2014; Hope et al., 2014).

In our study, we did not measure the time-line aspect of language processing *per se*; instead we used repetitive TMS to induce speech disturbances. We assumed that our rTMS train was delivered early enough (from 300 ms onwards) to be able to disturb semantic processing, phonological code retrieval, syllabification, phonetic encoding, and articulation components of the language processing (Indefrey and Levelt, 2004; Indefrey, 2011; Strijkers and Costa, 2011) needed in overt object and action naming. However, as speech processing is not only sequential but probably also occurs in parallel during several phases of the processing, the specific identification of the affected processes seems unreliable.

Recently, 300 and 0 ms nTMS pulse train onsets were compared to study the effects on sensitivity and specificity of picture naming during language mapping with navigated TMS. The 0-ms onset produced more specific results in the parietal areas when compared to DCS data (Krieg et al., 2014). The 0-ms paradigm

Table 2 | Single subject analysis within cortical areas.

Areas	Subjects (object/action)								Number of significant areas (sum over subjects)
	S1	S2	S3	S4	S5	S6	S7	S8	
anG	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	0/0
IFG	0.005/0.004	0.002/-	0.001/-	-/-	-/0.001	0.005/-	-/-	0/-	5/2
MFG	0.005/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	1/0
PoG	0.005/-	0.001/-	-/-	0.046/-	0.001/-	0/-	0.001/0.019	0/-	7/1
PrG	0.001/-	0.002/-	0.001/-	-/-	-/-	0.001/-	-/-	0.037/-	5/0
STG	-/-	-/-	0/0	0/-	0.010/-	-/-	0/0	0.005/-	5/2
MTG	-/-	-/-	0.002/-	-/-	-/-	-/-	0/-	-/-	2/0
SMG	0.001/-	0.001/0.004	-/-	-/-	0.033/-	0.018/-	0.023/-	-/-	5/1

Areas where nTMS induced statistically significant effects on object and action naming. The *p*-values are indicated; “-” means *p* > 0.05 and “0” values < 0.001. Notation: anG, angular gyrus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; PoG, postcentral gyrus; PrG, precentral gyrus; STG, superior temporal gyrus; MTG, middle temporal gyrus; SMG, supramarginal gyrus.

resembles more precisely the one applied in DCS so it is not surprising that the match between the 0-ms onset time and DCS is better. Nevertheless, the early onset of the nTMS may also influence conceptual preparation, lexical concept selection, and lemma retrieval attributed to the early stages of picture naming processes (Indefrey and Levelt, 2004; Indefrey, 2011). However, using the 300-ms latency for the nTMS pulse trains should have not biased our results to make object naming more sensitive to TMS than action naming, because action naming is a more demanding and time-taking process (e.g., Vigliocco et al., 2004; Mätzig et al., 2009).

Our results do not allow conclusions on cortical areas essential for processing of object-related or action-related words. Instead, they emphasize the network nature of language processing, in line with previous studies (Vigneau et al., 2006; Mätzig et al., 2009; Vigliocco et al., 2011; Lubrano et al., 2014). We did not confirm the previously described particular sensitivity of action naming for parietal cortical DCS (Corina et al., 2005). Our results were in line with more recent DCS results (Lubrano et al., 2014).

As the action naming was not specifically influenced by nTMS to posterior cortical areas, its use in preoperative speech mapping probably does not increase the sensitivity of nTMS in these regions. Therefore, the discordant results between nTMS and DCS of the posterior cortical areas (Picht et al., 2013; Tarapore et al., 2013), would probably not be improved by replacing the object naming with an action naming task.

Speakers name pictures of objects faster than those of actions, and action naming is more difficult than object naming in terms of accuracy and latencies (Vigliocco et al., 2002, 2004, 2011; Mätzig et al., 2009; Strijkers and Costa, 2011). This would suggest that action naming would be more easily disrupted by nTMS than object naming. However, the reverse was true in our experiment. Naming of words related with actions has been reported to involve more the motor cortex (Pulvermüller, 2005; Pulvermüller et al., 2005) and middle frontal gyrus (Lubrano et al., 2014). In our study, we did not stimulate those areas extensively enough to reach such conclusions. However, if the motor areas are more involved in action naming than object naming, it is possible that this “extra support” makes action naming less sensitive to TMS than object naming, when perisylvian regions are stimulated.

Object naming was particularly sensitive for nTMS to PoG, which is not typically studied by DCS when comparing object and action naming (Corina et al., 2005; Lubrano et al., 2014). However, in direct cortical recordings, spectral activity in PoG is modified during naming (Wu et al., 2011; Cogan et al., 2014). In fMRI, action naming induces stronger activation than object naming in PoG (Liljeström et al., 2008, 2009). It is possible that this stronger activation by action naming is less vulnerable to nTMS-induced perturbation.

Both fMRI and DCS have been used for language mapping. DCS during awake craniotomy is considered the gold standard for intraoperative brain mapping of cortical speech representations. However, it is demanding for the patient, strongly invasive, and may produce after-discharges, making the results difficult to interpret (Giussani et al., 2010). Moreover, the studied cortical regions are limited by the extent of craniotomy and demands of the surgery. Results from fMRI vary between different language paradigms and individuals, and its spatial accuracy in patients with gliomas has been questioned (Giussani et al., 2010; Wang et al., 2012). A recent case report suggests that nTMS may be more sensitive in defining speech lateralization than fMRI (Sollmann et al., 2013). The results of our study support the usefulness of picture naming combined with nTMS in presurgical planning (Krieg et al., 2013, 2014; Picht et al., 2013; Rösler et al., 2013). It also provides new possibilities for basic research of cortical speech representation. In addition, it may offer complementary information in comparison to other non-invasive methods (e.g., MEG and fMRI). Moreover, our results suggest that the efficacy of TMS in inducing naming errors can be modulated by the task; if a higher sensitivity is required, object naming is preferred; if a sparse amount of nTMS sites is required, action naming can be used.

Static pictures have limitations in exploring action naming performance, and some research groups have used videos of actions as stimuli to overcome this issue (e.g., Corina et al., 2005). Nevertheless, static images have been widely used in studies of action naming (see Mätzig et al., 2009). We did not match the frequency, familiarity length, or visual complexity of the pictures of objects vs. actions. Instead, subject-validated image stacks for objects and actions were obtained in the baseline naming session. It should be emphasized that we did not directly compare

the naming of objects vs. actions, but we compared the sensitivity of fluently named objects or actions to nTMS (for a similar approach, see Lubrano et al., 2014).

In summary, we have compared the naming error distributions induced by nTMS during object and action naming tasks. We suggest that object naming is more easily disrupted by nTMS than action naming. Particularly nTMS to PoG induced more errors during object naming than during action naming. Thus, use of action naming instead of object naming tasks most likely would not improve the specificity of nTMS in mapping posterior speech-related areas (Picht et al., 2013; Tarapore et al., 2013). These features, however, can be used in varying the sensitivity of functional mapping by nTMS for different cognitive paradigms in basic research and for presurgical planning. To resume, TMS applied to 8 subjects induced 93 errors during object naming and 33 during action naming. We find this surprisingly convincing for relatively small material, but believe that increasing the number of subjects will provide further important information for cortical speech organization.

ACKNOWLEDGMENTS

This study was financially supported by grants from the Helsinki University Hospital research fund and by the SalWe Research Program for Mind and Body (Tekes—the Finnish Funding Agency for Technology and Innovation grant 1104/10). We thank Dr. Thomas Picht, Department of Neurosurgery, Charité Hospital, Berlin, Germany, for providing us the action and object images, Antonios K. Thanellas for helping in the segmentation and registration of the MRIs to produce **Figure 2**, Dr. Ilkka Nissilä for his suggestions in the statistical analysis and the English examiner, Steve Lipson, M.A., for revising the English of our manuscript. Julio C. Hernandez-Pavon wants to thank CONACYT (Consejo Nacional de Ciencia y Tecnología) for funding.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2014.00660/abstract>

Supplementary Video | Clips from baseline recordings and from errors in object and action 32 naming during navigated TMS recorded with experimental setup 1.

REFERENCES

- Barker, A. T., Jalinous, R., and Freeston, I. L. (1985). Non-invasive magnetic stimulation of the human motor cortex. *Lancet* 1, 1106–1107. doi: 10.1016/S0140-6736(85)92413-4
- Berndt, R. S., Mitchum, C. C., Haendiges, A. N., and Sandson, J. (1997). Verb retrieval in aphasia. I. Characterizing single word impairments. *Brain Lang.* 56, 68–106. doi: 10.1006/brln.1997.1727
- Canessa, N., Borgo, F., Cappa, S. F., Perani, D., Falini, A., Buccino, G., et al. (2008). The different neural correlates of action and functional knowledge in semantic memory: an fMRI study. *Cereb. Cortex* 18, 740–751. doi: 10.1093/cercor/bhm110
- Cappelletti, M., Fregni, F., Shapiro, K., Pascual-Leone, A., and Caramazza, A. (2008). Processing nouns and verbs in the left frontal cortex: a transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 20, 707–720. doi: 10.1162/jocn.2008.20045
- Cogan, G. B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., and Pesaran, B. (2014). Sensory-motor transformations for speech occur bilaterally. *Nature* 507, 94–98. doi: 10.1038/nature12935
- Corina, D. P., Gibson, E. K., Martin, R., Poliakov, A., Brinkley, J., and Ojemann, G. A. (2005). Dissociation of action and object naming: evidence from cortical stimulation mapping. *Hum. Brain Mapp.* 24, 1–10. doi: 10.1002/hbm.20063
- Corina, D. P., Loudermilk, B. C., Detwiler, L., Martin, R. F., Brinkley, J. F., and Ojemann, G. (2010). Analysis of naming errors during cortical stimulation mapping: implications for models of language representation. *Brain Lang.* 115, 101–112. doi: 10.1016/j.bandl.2010.04.001
- Devlin, J. T., and Watkins, K. E. (2007). Stimulating language: insights from TMS. *Brain* 130, 610–622. doi: 10.1093/brain/awl331
- Epstein, C. M. (1998). Transcranial magnetic stimulation: language function. *J. Clin. Neurophysiol.* 15, 325–332. doi: 10.1097/00004691-199807000-00004
- Epstein, C. M., Lah, J. K., Meador, K., Weissman, J. D., Gaitan, L. E., and Dihenia, B. (1996). Optimum stimulus parameters for lateralized suppression of speech with magnetic brain stimulation. *Neurology* 47, 1590–1593. doi: 10.1212/WNL.47.6.1590
- Epstein, C. M., Meador, K. J., Loring, D. W., Wright, R. J., Weissman, J. D., Sheppard, S., et al. (1999). Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clin. Neurophysiol.* 110, 1073–1079. doi: 10.1016/S1388-2457(99)00047-4
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis. II. Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. doi: 10.1006/nimg.1998.0396
- Forster, M. T., Hattungen, E., Senft, C., Gasser, T., Seifert, V., and Szelényi, A. (2011). Navigated transcranial magnetic stimulation and functional magnetic resonance imaging: advanced adjuncts in preoperative planning for central region tumors. *Neurosurgery* 68, 1317–1325. doi: 10.1227/NEU.0b013e31820b528c
- Gerfo, E. L., Oliveri, M., Torriero, S., Salerno, S., Koch, G., and Caltagirone, C. (2008). The influence of rTMS over prefrontal and motor areas in a morphological task: grammatical vs. semantic effects. *Neuropsychologia* 46, 764–770. doi: 10.1016/j.neuropsychologia.2007.10.012
- Giussani, C., Roux, F. E., Ojemann, J., Sganzerla, E. P., Pirillo, D., and Papagno, C. (2010). Is preoperative functional magnetic resonance imaging reliable for language areas mapping in brain tumor surgery? Review of language functional magnetic resonance imaging and direct cortical stimulation correlation studies. *Neurosurgery* 66, 113–120. doi: 10.1227/01.NEU.0000360392.15450.C9
- Hagoort, P., and Indefrey, P. (2014). The neurobiology of language beyond single words. *Annu. Rev. Neurosci.* 37, 347–362. doi: 10.1146/annurev-neuro-071013-013847
- Hope, T. M., Prejawa, S., Parker Jones, Ö., Oberhuber, M., Seghier, M. L., Green, D. W., et al. (2014). Dissecting the functional anatomy of auditory word repetition. *Front. Hum. Neurosci.* 8:246. doi: 10.3389/fnhum.2014.00246
- Ilmoniemi, R. J., Ruohonen, J., and Karhu, J. (1999). Transcranial magnetic stimulation—A new tool for functional imaging of the brain. *Crit. Rev. Biomed. Eng.* 27, 241–284.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255. doi: 10.3389/fpsyg.2011.00255
- Indefrey, P., and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Kim, W. J., Min, Y. S., Yang, E. J., and Paik, N. J. (2013). Neuronavigated vs. Conventional repetitive transcranial magnetic stimulation method for virtual lesioning on the Broca's area. *Neuromodulation* 17, 16–21. doi: 10.1111/ner.12038
- Krieg, S. M., Shibani, E., Buchmann, N., Meyer, B., and Ringel, F. (2012). Presurgical navigated transcranial magnetic brain stimulation for recurrent gliomas in motor eloquent areas. *Clin. Neurophysiol.* 124, 522–527. doi: 10.1016/j.clinph.2012.08.011
- Krieg, S. M., Sollmann, N., Hauck, T., Ille, S., Foerschler, A., Meyer, B., et al. (2013). Functional language shift to the right hemisphere in patients with language-eloquent brain tumors. *PLoS ONE* 8:e75403. doi: 10.1371/journal.pone.0075403
- Krieg, S. M., Tarapore, P. E., Picht, T., Tanigawa, N., Houde, J., Sollmann, N., et al. (2014). Optimal timing of pulse onset for language mapping with navigated repetitive transcranial magnetic stimulation. *Neuroimage* 100, 219–236. doi: 10.1016/j.neuroimage.2014.06.016

- Liljeström, M., Hultén, A., Parkkonen, L., and Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Hum. Brain Mapp.* 30, 1845–1856. doi: 10.1002/hbm.20785
- Liljeström, M., Tarkiainen, A., Parvainen, T., Kujala, J., Numminen, J., Hiltunen, J., et al. (2008). Perceiving and naming actions and objects. *Neuroimage* 41, 1132–1141. doi: 10.1016/j.neuroimage.2008.03.016
- Lioumis, P., Zhdanov, A., Mäkelä, N., Lehtinen, H., Wilenius, J., Neuvonen, T., et al. (2012). A novel approach for documenting naming errors induced by navigated transcranial magnetic stimulation. *J. Neurosci. Methods* 204, 349–354. doi: 10.1016/j.jneumeth.2011.11.003
- Lubrano, V., Filleron, T., Démonet, J. F., and Roux, F. E. (2014). Anatomical correlates for category-specific naming of objects and actions: a brain stimulation mapping study. *Hum. Brain Mapp.* 35, 429–443. doi: 10.1002/hbm.22189
- Mätzig, S., Drunks, J., Masterson, J., and Vigliocco, G. (2009). Noun and verb differences in picture naming: past studies and new evidence. *Cortex* 45, 738–758. doi: 10.1016/j.cortex.2008.10.003
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., et al. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1293–1322. doi: 10.1098/rstb.2001.0915
- Noppeney, U., Josephs, O., Kiebel, S., Friston, K. J., and Price, C. J. (2005). Action selectivity in parietal and temporal cortex. *Brain Res. Cogn. Brain Res.* 25, 641–649. doi: 10.1016/j.cogbrainres.2005.08.017
- Pascual-Leone, A., Gates, J. R., and Dhuna, A. (1991). Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation. *Neurology* 41, 697–702. doi: 10.1212/WNL.41.5.697
- Petrovich Brennan, N. M., Whalen, S., de Moraes Branco, D., O'Shea, J. P., Norton, I. H., and Golby, A. J. (2007). Object naming is a more sensitive measure of speech localization than number counting: converging evidence from direct cortical stimulation and fMRI. *Neuroimage* 37, S100–S108. doi: 10.1016/j.neuroimage.2007.04.052
- Picht, T., Krieg, S. M., Sollmann, N., Rösler, J., Niraula, B., Neuvonen, T., et al. (2013). A comparison of language mapping by preoperative navigated transcranial magnetic stimulation and direct cortical stimulation during awake surgery. *Neurosurgery* 72, 808–819. doi: 10.1227/NEU.0b013e3182889e01
- Picht, T., Mularski, S., Kuehn, B., Vajkoczy, P., Kombos, T., and Suess, O. (2009). Navigated transcranial magnetic stimulation for preoperative functional diagnostics in brain tumor surgery. *Neurosurgery* 65, 93–98. doi: 10.1227/01.NEU.0000348009.22750.59
- Picht, T., Schmidt, S., Brandt, S., Frey, D., Hannula, H., Neuvonen, T., et al. (2011). Preoperative functional mapping for rolandic brain tumor surgery: comparison of navigated transcranial magnetic stimulation to direct cortical stimulation. *Neurosurgery* 69, 581–589. doi: 10.1227/NEU.0b013e31828181b9
- Poeppl, D., and Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition* 92, 1–12. doi: 10.1016/j.cognition.2003.11.001
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706
- Pulvermüller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797. doi: 10.1111/j.1460-9568.2005.03900.x
- Rösler, J., Niraula, B., Strack, V., Zdunczyk, A., Schilt, S., Savolainen, P., et al. (2013). Language mapping in healthy volunteers and brain tumor patients with a novel navigated TMS system: evidence of tumor-induced plasticity. *Clin. Neurophysiol.* 125, 526–536. doi: 10.1016/j.clinph.2013.08.015
- Ruohonen, J., and Karhu, J. (2010). Navigated transcranial magnetic stimulation. *Clin. Neurophysiol.* 40, 7–17. doi: 10.1016/j.neucli.2010.01.006
- Salmelin, R., Schnitzler, A., Schmitz, F., and Freund, H. J. (2000). Single word reading in developmental stutterers and fluent speakers. *Brain* 123, 1184–1202. doi: 10.1093/brain/123.6.1184
- Sanai, N., Mirzadeh, Z., and Berger, M. S. (2008). Functional outcome after language mapping for glioma resection. *N. Engl. J. Med.* 358, 18–27. doi: 10.1056/NEJMoa067819
- Siebner, H. R., Bergmann, T. O., Bestmann, S., Massimini, M., Johansen-Berg, H., Mochizuki, H., et al. (2009). Consensus paper: combining transcranial stimulation with neuroimaging. *Brain Stimul.* 2, 58–80. doi: 10.1016/j.brs.2008.11.002
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051
- Sollmann, N., Picht, T., Mäkelä, J. P., Meyer, B., Ringel, F., and Krieg, S. M. (2013). Navigated transcranial magnetic stimulation for preoperative language mapping in a patient with a left frontoopercular glioblastoma. *J. Neurosurg.* 118, 175–179. doi: 10.3171/2012.9.JNS121053
- Sörös, P., Cornelissen, K., Laine, M., and Salmelin, R. (2003). Naming actions and objects: cortical dynamics in healthy adults and in an anomic patient with a dissociation in action/object naming. *Neuroimage* 19, 1787–1801. doi: 10.1016/S1053-8119(03)00217-9
- Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. B* 64, 479–498. doi: 10.1111/1467-9868.00346
- Strijkers, K., and Costa, A. (2011). Riding the lexical speedway: a critical review on the time course of lexical selection in speech production. *Front. Psychol.* 2:356. doi: 10.3389/fpsyg.2011.00356
- Tarapore, P. E., Findlay, A. M., Honma, S. M., Mizuiru, D., Houde, J. F., Berger, M. S., et al. (2013). Language mapping with navigated repetitive TMS: proof of technique and validation. *Neuroimage* 82, 260–272. doi: 10.1016/j.neuroimage.2013.05.018
- Valyear, K. F., Cavina-Pratesi, C., Stiglick, A. J., and Culham, J. C. (2007). Does tool-related fMRI activity within the intraparietal sulcus reflect the plan to grasp? *Neuroimage* 36, T94–T108. doi: 10.1016/j.neuroimage.2007.03.031
- Vigliocco, G., Vinson, D. P., Damian, M. E., and Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition* 85, B61–B69. doi: 10.1016/S0010-0277(02)00107-5
- Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., and Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci. Biobehav. Rev.* 35, 407–426. doi: 10.1016/j.neubiorev.2010.04.007
- Vigliocco, G., Vinson, D. P., Lewis, W., and Garrett, M. F. (2004). Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cogn. Psychol.* 48, 422–488. doi: 10.1016/j.cogpsych.2003.09.001
- Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1435. doi: 10.1016/j.neuroimage.2005.11.002
- Vitikainen, A.-M., Lioumis, P., Paetau, R., Salli, E., Komssi, S., Metsähonkala, L., et al. (2009). Combined use of non-invasive techniques for improved functional localization for a selected group of epilepsy surgery candidates. *Neuroimage* 45, 342–348. doi: 10.1016/j.neuroimage.2008.12.026
- Wang, A., Peters, T. M., de Ribaupierre, S., and Mirsattari, S. M. (2012). Functional magnetic resonance imaging for language mapping in temporal lobe epilepsy. *Epilepsy Res. Treat.* 2012:198183. doi: 10.1155/2012/198183
- Whitaker, H. A., and Ojemann, G. A. (1977). Graded localisation of naming from electrical stimulation mapping of left cerebral cortex. *Nature* 270, 50–51. doi: 10.1038/270050a0
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45, S173–S186. doi: 10.1016/j.neuroimage.2008.10.055
- Wu, H. C., Nagasawa, T., Brown, E. C., Juhasz, C., Rothermel, R., Hoehstetter, K., et al. (2011). γ -oscillations modulated by picture naming and word reading: intracranial recording in epileptic patients. *Clin. Neurophysiol.* 122, 1929–1942. doi: 10.1016/j.clinph.2011.03.011

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 May 2014; accepted: 08 August 2014; published online: 02 September 2014.

Citation: Hernandez-Pavon JC, Mäkelä N, Lehtinen H, Lioumis P and Mäkelä JP (2014) Effects of navigated TMS on object and action naming. *Front. Hum. Neurosci.* 8:660. doi: 10.3389/fnhum.2014.00660

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Hernandez-Pavon, Mäkelä, Lehtinen, Lioumis and Mäkelä. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity

Keke Yu¹, Ruiming Wang^{1*}, Li Li² and Ping Li^{3*}

¹ Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, China

² College of International Culture, South China Normal University, Guangzhou, China

³ Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, Pennsylvania, PA, USA

Edited by:

Patti Adank, University College
London, UK

Reviewed by:

Nikhil Sharma, University College
London, UK

Xiaoqing Li, Institute of Psychology,
Chinese Academy of Sciences,
China

*Correspondence:

Ruiming Wang, Center for Studies
of Psychological Application, School
of Psychology, South China Normal
University, No. 55, West Zhongshan
Ave., Tianhe District, Guangzhou
510631, China
e-mail: wangrm@scnu.edu.cn;
Ping Li, Department of Psychology
and Center for Brain, Behavior, and
Cognition, Pennsylvania State
University, University Park, 201 Old
Main, Pennsylvania, PA 16802, USA
e-mail: pul8@psu.edu

The accurate perception of lexical tones in tonal languages involves the processing of both acoustic information and phonological information carried by the tonal signal. In this study we evaluated the relative role of the two types of information in native Chinese speaker's processing of tones at a preattentive stage with event-related potentials (ERPs), particularly the mismatch negativity (MMN). Specifically, we distinguished the acoustic from the phonological information by manipulating phonological category and acoustic interval of the stimulus materials. We found a significant main effect of phonological category for the peak latency of MMN, but a main effect of both phonological category and acoustic interval for the mean amplitude of MMN. The results indicated that the two types of information, acoustic and phonological, play different roles in the processing of Chinese lexical tones: acoustic information only impacts the extent of tonal processing, while phonological information affects both the extent and the time course of tonal processing. Implications of these findings are discussed in light of neurocognitive processes of phonological processing.

Keywords: Chinese lexical tones, acoustic processing, phonological processing, mismatch negativity (MMN), preattentive stage

INTRODUCTION

The use of lexical tones to differentiate lexical semantics is a characteristic of Chinese and other tonal languages. According to Yip (2002), the world's languages can be categorized into three types depending on the role that the pitch information plays in the expression of meaning: tone language (e.g., Chinese), intonation language (e.g., English), and pitch-accent language (e.g., Japanese). In languages like Chinese, different lexical semantics are expressed through the variations of pitch height and pitch contour at the syllable level, as opposed to intonation languages where pitch variations occur only at the phrase or sentence level, or pitch-accent language where pitch variations occur between syllables. In recent years, researchers have become interested in the neurocognitive processes associated with tonal languages (see Gandour, 2006; Jongman et al., 2006 for reviews). Several studies have also used neuroimaging techniques including event-related potential (ERP) and functional magnetic resonance imaging (fMRI) to study the processing of lexical tones in Chinese (e.g., Gandour et al., 2000; Klein et al., 2001; Li et al., 2008; Zhang et al., 2011; Wang et al., 2013a).

When a native speaker processes lexical tones in Chinese,¹ the speaker deals with at least two types of information: the acoustic information that includes the physical features of auditory input such as the fundamental frequency (F0), and the phonological information that expresses lexical semantics on the basis of which word categories are identified (Luo et al., 2006; Xi et al., 2010). To investigate the difference between the processes of the acoustic vs. the phonological information in lexical tones, one line of previous research has been to understand whether lexical tonal processing in Chinese is left or right lateralized in the brain. The functional hypothesis claims that when tones are processed as phonological units they are lateralized to the left hemisphere, whereas when they are processed as purely acoustic information they are lateralized to the right hemisphere (van Lancker, 1980; Wong, 2002). Chinese lexical tones contain both the acoustic information and the phonological information.

¹Although our experimental materials involve lexical tones only from Mandarin Chinese, the findings from this study apply to all Chinese dialects. We therefore use "Chinese" as a generic term henceforth in the paper to refer to all Chinese dialects.

Thus, based on this hypothesis, Chinese tonal processing involves both hemispheres for different processing of these two kinds of information. A competing theory, the acoustic hypothesis, claims that the brain lateralization of tonal processing depends on the acoustic properties of the auditory input. Spectral variations such as those contained in pitch information are preferentially processed in the right hemisphere, whereas temporal variations such as those contained in vowels are processed more strongly by the left hemisphere (Zatorre and Belin, 2001; Zatorre et al., 2002). According to this hypothesis, the acoustic contrasts and the phonological contrasts in Chinese tones are both spectral variations, and so both acoustic processing and the phonological processing are lateralized to the right hemisphere.

In light of these views on brain lateralization, Gandour et al. have put forward a comprehensive hypothesis on this issue (Gandour et al., 2000, 2004; Gandour, 2006). They suggested that lexical tonal processing engages both hemispheres, depending on the types of information involved during processing. In particular, for the same lexical tones, the left hemisphere is more involved in the semantic processing whereas the right hemisphere more in the acoustic processing of the pitch information. Given that processing of lexical tones involves the processing of both acoustic information and phonological information (which is semantically differentiating), this hypothesis suggests that there is no simple lateralization pattern associated with the brain's processing of tones.

The key issue under consideration here is therefore whether processing involves acoustic features of tones or phonological features of tones. Luo et al. (2006) used mismatch negativity (MMN) to study the features of tonal processing. MMN is a powerful method to examine the early stages of acoustic and phonological processing, as previous studies have indicated (e.g., Näätänen et al., 1978, 2007; Näätänen and Alho, 1997). It is an electrophysiological component that reflects the brain's automatic detection of deviant patterns that do not match the general pattern that has been observed, which peaks at typically around 200–250 ms after stimulus onset, mostly in the frontocentral areas. MMN responses can be elicited by an oddball paradigm in which several infrequent deviant stimuli are embedded in frequent standard stimuli during auditory presentation. Luo et al. (2006) found that the mean amplitude of MMN elicited by tone deviants in the right hemisphere was larger than that in the left hemisphere. Their conclusion was that at the pre-attentive stage, a stage that is characterized by automatic processing when people are unconscious about the detailed properties of the stimuli (Kubovy et al., 1999), listeners mainly process the acoustic information of lexical tones. At this stage, the processing is usually lateralized to the right hemisphere. At a second stage of tone processing, the attentive stage, a stage in which people process the presented stimuli consciously, listeners tended to process the semantic information of tones via the left hemisphere. Luo et al.'s novel finding was that, in addition to the work of Gandour et al. whether the cognitive processing of Chinese lexical tones involves both hemispheres will depend on the different processing stages. Thus, the MMN provides an excellent measurement of the time course of processing, contributing to additional insights in lexical tone processing.

In addition to the lateralization debate of tone processing in Chinese, recent studies have also highlighted the role of categorical perception of tones. Categorical perception refers to the ability that human listeners can perceive continuous acoustic signals as discrete linguistic representations: listeners are sensitive to the boundaries between different phonetic categories, but are insensitive to acoustic changes within the boundaries of same phonetic category (Liberman et al., 1957, 1967). This across-category vs. within-category perception difference has been extensively studied in previous work with segmental phonemes, such as consonants and vowels (e.g., with VOT characteristics), but are less well understood in studies with suprasegmental features such as tones. It has been found recently, however, that the perception of lexical tones shows categorical perception just as do phonemes: native speakers of tonal languages are more sensitive to across-category tonal variations than within-category variations (see Francis et al., 2003; Hallé et al., 2004; Xu et al., 2006; Xi et al., 2010).

Xi et al. (2010) used MMN to examine categorical perception of tones in Chinese. The authors found that although both across-category and within-category variations of tones elicited MMNs in bilateral frontal-central areas, the former elicited larger MMNs in the left than that in the right hemisphere, whereas the latter elicited larger MMNs in the right than in the left hemisphere. These patterns provide support for the categorical perception of tones, while at the same time evidence for the hypothesis that both hemispheres are involved in the processing of tones (Gandour, 2006), in that the processing of within-category stimuli mainly involves the acoustic processing of pitch information while the processing of across-category stimuli the phonological information. Moreover, within the time window of MMN (200–250 ms), both acoustic information and phonological information are processed in parallel, which contrasts to the two-stage hypothesis of Luo et al. (2006).

An fMRI study by Zhang et al. (2011) further indicated that the interaction between acoustic processing and phonological processing in the two hemispheres: across-category variations elicited stronger activation in the left middle temporal gyrus than did the within-category variations, whereas within-category variations elicited stronger activation in the right superior temporal gyrus. These fMRI findings indicate how low-level acoustic analysis (within-category) is modulated by high-level phonological representations (across-category), and are therefore consistent and complementary with the MMN findings reported in Xi et al. (2010). In another ERP study of categorical perception of tones at the attentive stage, Zhang et al. (2012a) found that the conscious processing of the within-category stimuli and the across-category stimuli involved the N2b and P3b components. The data were compatible with the response patterns as Xi et al. (2010): for both N2b and P3b, across-category stimuli elicited larger response in the left recording sites than the right; while the within-category stimuli elicited the same response in both hemispheres.

Such findings of categorical perception of tones also have significant implications for understanding disorders during processing. For example, Zhang et al. (2012b) found that Chinese-speaking children who were at risk for dyslexia

showed no significant differences between the across-category vs. within-category stimuli, in contrast to both adults and age-matched normally developing children. It could be that children with reading disorders may perceive the phonological information in the same way as they do with acoustic information.

One important question that remains unclear from previous research is whether acoustic information and phonological information of tones are fundamentally different (i.e., different kinds of information), or whether they are only different on a continuum (same kinds of information). As suggested by the functional hypothesis (van Lancker, 1980), the pitch information contained in Chinese lexical tones can be divided into two types depending on whether it serves as acoustic signal or phonological unit: these are two distinct kinds and also have different cognitive processing consequences. Gandour et al. (2000) further distinguished between the pure acoustic features vs. acoustic features plus semantic features in the perception of tones such as lexical tones in Chinese and Thai. An alternative view to the above is that there is no fundamental difference between the two in terms of cognitive processing: the pitch contrasts in tones are spectral variations, and the processing of these contrasts always involves the same type of acoustic analysis regardless of its specific features (e.g., Zatorre and Belin, 2001; Warrier and Zatorre, 2004; Ren et al., 2009). This alternative view suggests that the fundamental features of the acoustic information and the phonological information both lie in the spectral variations, and therefore are similar with regard to neural mechanisms, the processing of phonological information and that of acoustic information involve the same neural patterns, both in the right hemisphere. This argument contrasts with the suggestion of different neural correlates of acoustic vs. phonological processing, as discussed above (e.g., MMN and fMRI evidence of left vs. right lateralized patterns).

In the present study, we designed an experiment to further explore the different roles that acoustic vs. phonological information may play in the perception of Chinese tones. In particular, we relied on the MMN patterns that have been previously used successfully in this area (Luo et al., 2006; Xi et al., 2010). Previous work, including Xi et al. (2010) and Zhang et al. (2011), has examined only the contrast between acoustic information and phonological information as the contrast between within-category and across-category variations. In the MMN paradigm, this means that the across- or within-category deviants are equally spaced when compared with the standard stimuli of tonal contrasts. In the current study, to further differentiate the two types of information, we added a new variable, the acoustic interval, which refers to the different F0 interval between the deviant and the standard stimuli. This new variable has two levels, small acoustic interval vs. large acoustic interval.

The stimuli materials used by Xi et al. (2010) and Zhang et al. (2011) were chosen from a Chinese lexical tonal continuum, from the high-rising tone (tone 2, /pa2/) to the falling tone (tone 4, /pa4/). The differences between the two tones can be acoustically manipulated according to the F0 contour, to produce a range of in-between tones, resulting in a continuum of stimulus 1 to stimulus 11 (see Figure 1 and Section Materials and Methods for details). Previous studies were focused on categorical

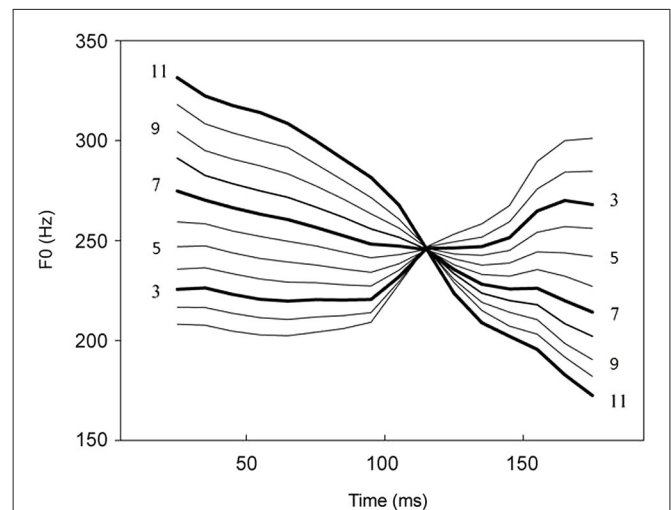


FIGURE 1 | Tone contours of the continuum from /ba2/ to /ba4/ (from Xi et al., 2010; reproduced with permission from authors and publisher). Stimuli 3, 5, 7, 9 and 11 are marked. Continua 3, 7 and 11 marked with thick lines were used in previous study (Xi et al., 2010; Zhang et al., 2011, 2012a,b). In the present study, continua 5, 9 were also chosen to test the acoustic interval variable.

perception of tones, so only stimuli 3, 7, 11 were used in their experiments (e.g., Xi et al., 2010; Zhang et al., 2011). In the current study, we continued to use the 11 stimulus set, but we included stimuli 5 and 9 in our experiment, to produce an orthogonal design that involved both phonological categories and acoustic intervals (see Section Materials under Methods for details). This design would allow us to systematically test the role of acoustic vs. phonological information in the processing of Chinese lexical tones, with both small and large interval differences in the acoustic signal and across and within differences in the phonological category. Specifically, we predict that the MMN patterns may reflect different impact of these two types of information, with regard to both the magnitude and the peak latency. We hypothesize that if the acoustic and the phonological information belongs to the same type of information (different dimensions of the auditory input), as proposed by Zatorre et al. the variations in the MMN mean amplitudes and peak latencies would be similar. On the other hand, if these two types of information are fundamentally different and reflect different cognitive processes, we may see variations of MMN patterns as a function of their differences, for example, in amplitude and time course.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-six neurologically healthy volunteers (21 females, mean age 20 years, range 19–21 years) took part in the study. All participants had normal hearing and minimal musical experience, and were native speakers of Mandarin Chinese recruited from the South China Normal University. The participants were all right-handed basing on their self-assessments. They gave written consent before they took part in the experiment and received monetary compensation for taking part in the experiment.

This study was approved by the ethics review board of South China Normal University.

EXPERIMENTAL DESIGN

Our study used a within-subject design with two factors: phonological category (within-category/ across-category), and acoustic interval (large/small). See *Materials* section below for details. The dependent variables were the mean amplitude and the peak latency of the MMNs elicited by the stimuli.

MATERIALS

The materials were the same as used in Xi et al. (2010; see **Figure 1**), except noted below. They were acoustically manipulated based on Chinese lexical tones, to produce a continuum from the high-rising tone (tone 2) to the falling tone (tone 4). Two Chinese monosyllables, /pa2/ and /pa4/, were firstly produced by a native speaker and digitally edited by Sound-Forge (SoundForge9, Sony Corporation, Japan) to get a same duration (200 ms). Then the two monosyllables were further edited by the Praat software² to keep the same acoustic features except the pitch contour. They were then used as the endpoint stimuli to create the 10-interval lexical tone continuum in Matlab (Mathworks Corporation, USA) using the toolbox of STRAIGHT (Kawahara et al., 1999). Thus there were 11 artificially generated stimuli in the continuum that differed only in the F0 (labeled as stimulus 1 to stimulus 11). The F0 intervals between any two adjacent stimuli were also acoustically manipulated to be the same, for the purpose of the experiment.

In Xi et al. (2010) study, stimuli 3, 7, and 11 were used to construct the across-category pair (3 vs. 7) and the within-category pair (7 vs. 11) in their experiment. These pairs of contrast between 3 and 7, and 7 and 11, had large acoustic intervals for both across- and within-category contrasts. Given the focus of the previous studies on categorical perception, they served to address the research questions well. In the present study, we constructed an orthogonal design incorporating two variables, phonological category (as in Xi et al., 2010), and acoustic interval, in the following way: stimuli 5 vs. 7 (small interval, across category), stimuli 3 vs. 7 (large interval, across category), 7 vs. 9 (small interval, within category), stimuli 7 vs. 11 (large interval, within category). This new design would allow us to address additional research questions, as discussed in the Introduction section.

To ensure that the materials meet the need of our ERPs experiment, we conducted a norming study for the experimental material with a separate group of participants. 15 participants took part in an identification task and a discrimination task. In the identification task, they were asked to identify whether each stimulus (stimulus 3, 5, 7, 9, 11) was the high-rising tone (tone 2) or the falling tone (tone 4). 20 trials for each stimulus were randomly presented in isolation, with no stimulus being presented consecutively three times. In the discrimination task, participants were asked to judge whether a pair of presented stimuli were the same or not. The stimulus pairs were presented randomly, and the pairs which contained different stimuli were presented in both directions, each for 10 trials. To balance the

number of “yes” or “no” responses, we included pairs of the same stimulus, each presented for 16 times (80 in total). Practice trials were given to the participants before the actual norming experiment.

PROCEDURE

An improved passive oddball paradigm, proposed by Duncan et al. (2009), was used for this experiment. Classic oddball paradigms usually contained one kind of deviant stimuli and one kind of standard stimuli in a block, whereas the improved paradigm contained more than one kind of deviant stimuli in a block. The improved oddball paradigm has been shown to produce MMNs more quickly and effectively. A total of 1015 stimuli were presented to the participants during the experiment, including 15 standard stimuli at the beginning of the experiment and followed by 600 standard stimuli and 400 deviant stimuli, each type of deviant being 100. 15 standard stimuli were firstly presented to promote the participants to adapt to the experiment. The deviants were presented pseudo-randomly among standards, and any two adjacent deviants were different. Each stimulus was presented for 200 ms. The stimulus-onset-asynchrony (SOA) was set to be 800 ms. The stimulus presentation lasted for about 15 min.

Participants took a passive auditory task. They were instructed to watch a movie (*Whisper of the Heart*) (e.g., Duncan et al., 2009) in which the sound tracks of the movie were turned off. Moreover, they saw the movie for 5 min first before starting the experiment. No overt responses were required of the participants. To ensure that the participants paid attention to the movie, the experimenter asked them to answer five questions about the movie after the experiment. The experiment lasted 22 min in total (5 min in movie plus 15 min in experimental session proper).

ELECTROENCEPHALOGRAPH (EEG) RECORDING

EEG was recorded using a 64-channel (Ag–AgCl) NeuroScan system (NeuroScan). Electrodes were positioned following the 10–20 system convention. The reference electrode was placed at the tip of the nose and the ground electrode was placed at FPz. The vertical electro-oculogram (EOG) was recorded supra- and infra-orbitally from the left eye. The horizontal EOG was recorded as the left vs. right orbital rim. The impedance of each electrode was kept below 5 k Ω . EEG and EOG signals were digitized online at 500 Hz and band-pass filtered from 0.05 to 100 Hz.

DATA ANALYSIS

Off-line signal processing was carried out using Scan 4.3 (NeuroScan). The reference electrode was converted first to bilateral mastoid (M1 and M2) and some artifacts were rejected manually. Data from two participants were excluded from further analyses due to their excessive eye blinking. Data were then adjusted by eliminating the interference of the horizontal and vertical eye-movements. The data were segmented for a 700 ms time window, including a 100-ms pre-stimulus baseline. Then the baseline was corrected according to Zhao (2010) and the recorded trials with eye blinks or other activities beyond the range of -80 to 80 mV were rejected. The data from the whole-head recordings were off-line band-pass filtered (1–30 Hz) with a finite impulse

²<http://www.fon.hum.uva.nl/praat/>

response filter. Finally, the ERPs evoked by the standard stimuli and the deviant stimuli were calculated by taking the averages of individual trials from each subject. Only those data with at least 80 accepted deviant trials in each deviant condition were adopted. With this criterion, data from another two participants were excluded from further analyses. MMNs were then derived by subtracting the ERPs evoked by the standard stimuli from those evoked by the deviant stimuli.

On the basis of findings from previous studies in the literature, we selected three recording sites for statistical analyses: F3, F4 and FZ. In previous studies of MMN, the MMN component typically peaks around 200–250 ms (Näätänen et al., 1978, 2007; Näätänen and Alho, 1997). Researchers usually choose a time window around this peak based on the grand-average waveforms of their particular data, which could include 100–350 ms (e.g., Kaan et al., 2008), 150–300 ms (e.g., Tsang et al., 2011), 230–360 ms (Wang et al., 2013b), and so on. Considering the usual range of MMN's peak and the grand-average waveforms of the present study, we chose a time window of the MMNs to be 200–350 ms. The peak of MMN for each subject at the four conditions in this time window was detected by using the procedure “Peak detection” in Scan 4.3 (NeuroScan). The MMN mean amplitudes were calculated by averaging the responses within the time window ranging from 20 ms before the peak of MMN recorded from electrode FZ to 20 ms after that peak. The mean amplitudes and peak latencies of the three chosen recording sites (F3, F4, FZ) were used for further statistical analyses.

RESULTS

NORMING EXPERIMENT

In the identification task, participants were asked to identify whether each stimulus (stimulus 3, 5, 7, 9, 11) was the high-rising tone (tone 2) or the falling tone (tone 4). In the discrimination task, participants were asked to judge whether a pair of presented stimuli were the same or not. We analyzed the proportions of different tone judgments in the identification task and the proportions of “yes” or “no” responses in the discrimination task

according to Xi et al. (2010) study. In the identification task, the proportion of different stimuli, 3, 5, 7, 9, 11, which were regarded as tone 4, was 4.4%, 8%, 87.8%, 97.4%, 98.2%, respectively. These results showed that the participants identified stimuli 3 and 5 as tone 2 and stimuli 7, 9, 11 as tone 4 indeed. In the discrimination task, the proportion of negative judgments for pairs 3–7, 5–7, 9–7, 11–7 was 89.7%, 89.7%, 15%, 30%, respectively. These results showed that the participants identified stimuli 3 and 7, and 5 and 7 as different tones, and stimuli 7 and 9, and 7 and 11 as the same tone.

The norming experiment showed that we can reliably treat stimuli 3 and 7, 5 and 7 as across-category pairs and stimuli 7 and 9, and 7 and 11 as within-category pairs. Considering the F0 intervals as illustrated above, stimuli 3 and 7, 11 and 7 were treated as the large interval stimulus pairs, and the F0 intervals of each pair were four F0 units. Similarly, stimuli 5 and 7, 9 and 7 were treated as the small interval stimulus pairs, and the F0 intervals of each pair were two F0 units. These results helped us to determine stimulus 7 in the tonal continuum (**Figure 1**) as the standard stimulus, stimulus 3 as an across-category with large interval deviant, stimulus 5 as an across-category with small interval deviant, stimulus 9 as a within-category with small interval deviant, and stimulus 11 as a within-category with large interval deviant.

ERP EXPERIMENT

Figure 2 presents the grand average waveforms of the ERPs elicited by the standard stimulus and four deviants at F3, FZ, F4 electrode locations. As shown in **Figure 2**, different waveforms of the ERPs to the standard stimuli and the four deviants were observed at the three electrode locations.

To ensure that the deviant stimuli elicited MMNs, we conducted four paired samples *t*-tests to compare the grand average amplitudes of the four deviant contrasts in the time window of MMN at three electrode locations. The difference between the grand average amplitudes of stimuli 3, 5, 11 and standard stimuli were significant ($t_{(1,31)} = 6.124, p < 0.001$; $t_{(1,31)} = 3.928$,

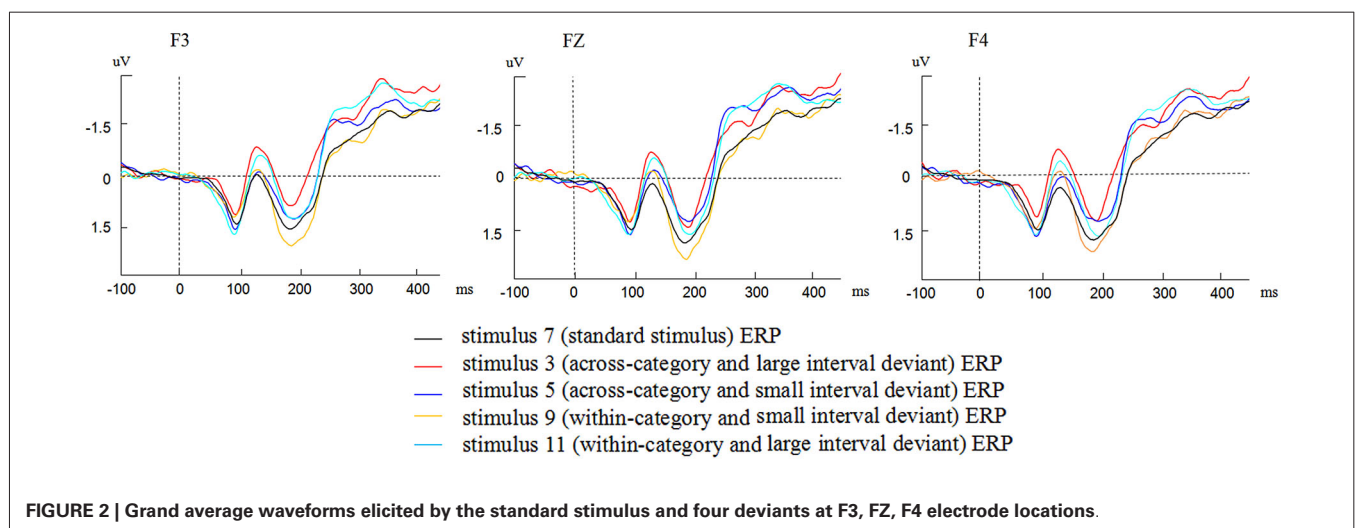


FIGURE 2 | Grand average waveforms elicited by the standard stimulus and four deviants at F3, FZ, F4 electrode locations.

$p < 0.001$; $t_{(1,31)} = 2.625$, $p = 0.013 < 0.05$). The difference between stimuli 9 and standard stimuli was marginally significant ($t_{(1,31)} = 1.718$, $p = 0.096 < 0.1$). Thus the four deviant stimuli elicited reliable MMNs.

Figure 3 presents the different MMNs obtained by subtracting the ERP waveforms of the standard stimuli from those of the deviant stimuli at F3, FZ, F4 electrode locations. In the figure, distinct MMNs were displayed in the MMN time window (200–350 ms) at the three electrode locations.

We conducted two separate repeated-measures 2×2 ANOVAs, with phonological category (within-category/across-category) and acoustic interval (large/small) as independent variables, one for the mean amplitude and another for the peak latency of the MMNs. For all analyses, degrees of freedom were adjusted according to the Greenhouse–Geisser method when appropriate.

MMN peak latency

Figure 4 presents the mean peak latencies of MMNs at F3, F4, FZ electrodes (the line segment represents one standard error). ANOVA results showed a significant main effect of phonological category ($F_{(1,31)} = 15.256$, $p < 0.001$, across-category < within-category), reflecting that the across-category deviants were processed earlier than the within-category deviants. There was no significant main effect of acoustic interval, nor an interaction between phonological category and acoustic interval ($ps > 0.10$).

MMN mean amplitudes

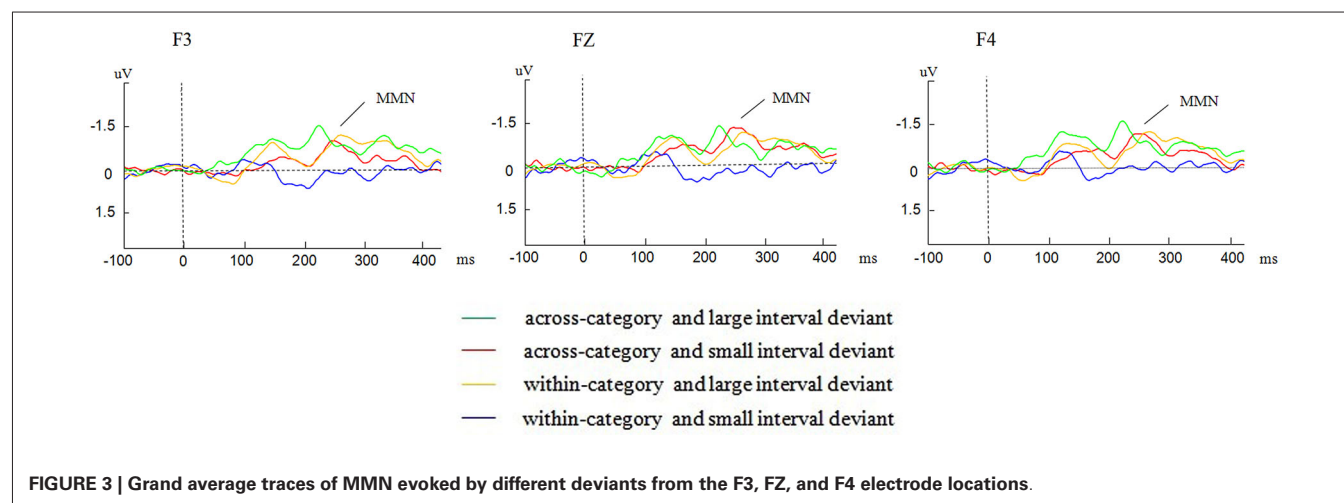
Figure 5 presents the mean peak amplitudes of MMNs at F3, F4, FZ electrodes. The main effect of phonological category was significant ($F_{(1,31)} = 20.312$, $p < 0.001$, across-category > within-category), showing that the processing of across-category deviants elicited greater MMN patterns than the processing of within-category deviants. There was also a significant main effect of acoustic interval ($F_{(1,31)} = 9.924$, $p < 0.01$, large interval > small interval), indicating that the processing of large interval deviants elicited greater MMN patterns than the processing of small interval deviants. There was no significant interaction between phonological category and acoustic interval ($p > 0.10$).

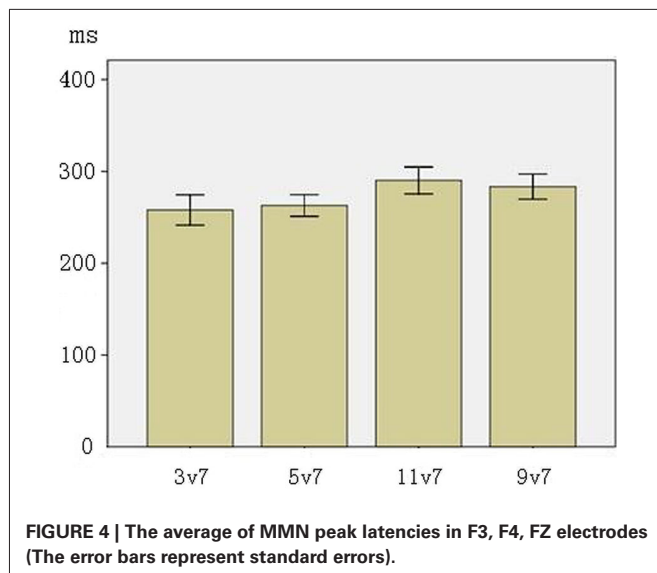
DISCUSSION

The perception of lexical tones in Chinese consists of the processing of the acoustic and the phonological information. But it has been debatable in the literature whether these two types of information are fundamentally different or whether they belong to the same auditory input that involves similar cognitive processes. As we reviewed earlier in the Introduction section, previous studies have investigated the differences between the acoustic information and the phonological information from the perspective of brain lateralization (Xi et al., 2010; Zhang et al., 2011, 2012a). The present study has focused on the different roles of acoustic vs. phonological information in native Chinese listener's processing through acoustic interval and phonological category. In particular, we used a paradigm from the MMN studies of categorical perception of Chinese lexical tones, examining four kinds of stimuli with different phonological categories (across vs. within) and distinct acoustic intervals (large vs. small).

The results from this study show that the acoustic and the phonological information of Chinese lexical tones have distinct impacts on processing as revealed by MMNs. The main effect of phonological category and acoustic interval for MMN mean amplitudes are both statistically significant. In particular, MMNs elicited by across-category deviants were larger than those by within-category deviants and MMNs in response to large interval deviants were greater than small interval deviants. The results revealed that both types of information influenced MMN mean amplitude. But only phonological category showed a significant main effect for MMN peak latency: the MMN peak latencies for the across-category deviants were shorter than for the within-category deviants.

In the ERP literature, the mean amplitude of ERP components has been used to reflect the extent of neural resources during cognitive processing, and the peak latency has been argued to indicate the time course of the process (e.g., Duncan et al., 2009). The present study found that the acoustic information with different intervals only affected the MMN mean amplitude, but the phonological information influenced both the mean amplitude and the peak latency of the MMNs. The results revealed that





the acoustic information of Chinese lexical tones impacted the extent of tonal processing while the phonological information affected not only the extent but also the time course of tonal processing. These patterns showed the different roles that the acoustic information and the phonological information play in the processing of Chinese lexical tones at the pre-attentive stage.

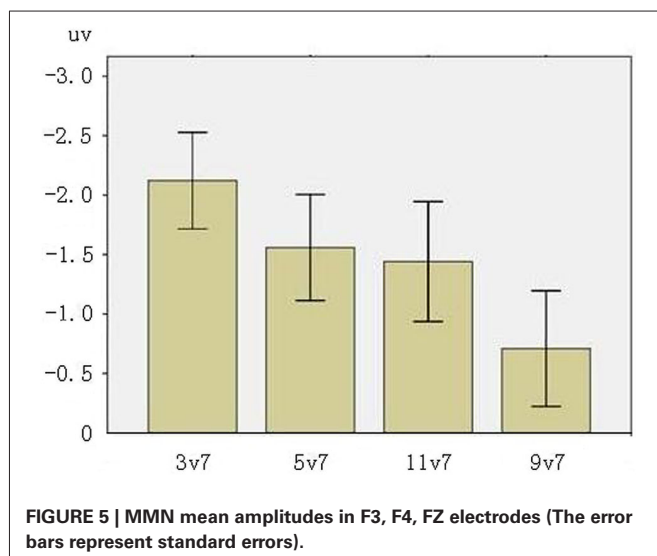
As discussed earlier, Gandour et al. have argued for the functional hypothesis of tonal processing, according to which the phonological information contained in pitch information, such as in tones, is different from the general acoustic information, because the former expresses lexical semantic differences (e.g., Gandour et al., 2000, 2002; Wong, 2002). An alternative view, the acoustic hypothesis proposed by Zatorre et al. perceives the pitch information in tones as no different from general acoustic information (e.g., Zatorre and Belin, 2001; Warrier and Zatorre, 2004). In the present study, distinct profiles in MMN mean amplitudes and peak latencies were observed to be associated

with the acoustic and the phonological variations. These results provide support to the functional hypothesis rather than the acoustic hypothesis.

Some previous studies such as Luo et al. (2006), proposed that acoustic information of tones is mainly processed at the pre-attentive stage while phonological information is processed at the attentive stage. The proposal agreed with the view that the processing of the acoustic information and the phonological information are different. However, this proposal emphasized that the two kinds of information are processed at different stages. It is generally agreed that people can automatically process the stimuli without attention at the pre-attentive stage. Recent studies, in contrast to Luo et al. (2006), have suggested that acoustic and phonological information may be processed at both attentive and pre-attentive stages in parallel (Xi et al., 2010; Zhang et al., 2012b). In the present study, we differentiated the acoustic and the phonological information with a more fine-grained level, especially with regard to the acoustic interval of the stimulus. That is, the stimuli contained not only phonological category differences (different tones), but also acoustic interval changes (same tone with different F0). The results showed that both the acoustic and the phonological information of Chinese lexical tones can be processed at pre-attentive stage by native listeners. In addition, the present study found that the peak latencies of MMNs elicited by the across-category deviants were earlier than by the within-category deviants. Thus, phonological information associated with lexical tones may be processed even before the acoustic information at pre-attentive stage. This is a rather surprising finding that directly contrasts with arguments of the two-stage model by Luo et al. (2006).

How do we account for the surprising finding that phonological information may be activated even earlier than acoustic information in Chinese lexical tone processing? As discussed above, MMN reflects the automatic detection of distinct stimuli, acoustic or phonological, by the human brain. Considering the auditory stimuli in the present study, the within-category stimuli differed only in the acoustic information, while the across-category stimuli involved differences in both the phonological and the acoustic dimension. Thus, MMN may reflect the latter differences more easily and quickly, which is why we observed shorter peak latency of the across-category stimuli.

Some researches considered that the MMNs elicited by tone variations may result from the long-term memory trace of Chinese lexical tones. Chandrasekaran et al. (2007a) assessed the different MMN responses to non-speech stimuli involving similar pitch variations as Chinese lexical tones by Chinese and English listeners. Their results showed that Chinese listeners had larger MMN responses than English listeners, which indicated that the same non-speech stimuli involving pitch variations may activate long-term memory traces of lexical tones for Mandarin listeners (Chandrasekaran et al., 2007a). The patterns from the present study showed that the MMNs elicited by the stimuli involving both semantic variations and acoustic variations were larger and earlier than the stimuli only involving acoustic variations in mean amplitude. It seems that long-term memory representation of tones may have contributed to these patterns. Thus, based on Chandrasekaran et al.'s finding and our own results, the activation



of the long-term phonological memory trace may play a significant role to enhance tone perception with regard to the processing of phonological information.

Pitch contour refers to the direction of change in F0 according to Gandour (1983). Chandrasekaran et al. (2007b) investigated whether different pitch contours of Chinese lexical tones have different impacts on tonal processing by tone 1, tone 2 and tone 3. Their results showed that the MMN peak latency elicited by tone 1 vs. tone 3 was earlier than that elicited by tone 2 vs. tone 3. As the difference between tone 1 and tone 3 was larger than that between tone 2 and tone 3 in pitch contour shapes, they concluded that the different pitch contour shapes impacted the MMN peak latencies. In the present study, we compared the within-category stimuli (9 vs. 7; 11 vs. 7) with the across-category stimuli (3 vs. 7; 5 vs. 7), which have different direction of pitch contour (see **Figure 1**). That is, the difference between the across-category stimuli and the standard stimuli was larger than that between the within-category stimuli and the standard stimuli in pitch contour shapes. Our result showed that the peak latency of across-category deviants was earlier than that of within-category deviants. This pattern is consistent with Chandrasekaran et al. (2007b) and further indicated the impact that pitch contour shapes have on the MMN peak latency. In addition, as discussed earlier, the present study further indicated the interaction of two independent variables, showing the different impacts of acoustic interval and phonological category in lexical tone processing. However, there remains the question of whether the changes in pitch contour shapes and the changes in phonological information have the same amount of impact on tonal processing. Our current study design cannot yet address this question and it should be examined in future research.

In conclusion, our study has explored MMN patterns (mean amplitude and peak latency) to identify the cognitive processes associated with the pre-attentive stage processing of both acoustic and phonological information involved in the perception of Chinese lexical tones. The results showed that the acoustic information of tones only impacts the extent of tonal processing, whereas the phonological information impacts the time course as well as the extent of the processing. Our data suggest that the acoustic information and the phonological information of tones were distinct auditory inputs, at least to the native listeners of Chinese who have long-term experience with the representation of pitch information that differentiates lexical meanings.

ACKNOWLEDGMENTS

We thank Hua Shu, Linjun Zhang, and Yang Zhang for providing the experimental materials and helping to experimental design. This work was supported by the National Natural Science Foundation of China (31200785), National Social Science Foundation of China (11CYY023, CBA130125), and Graduate Research Innovation Foundation of South China Normal University (2013kyjj083).

REFERENCES

- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2007a). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport* 18, 1963–1967. doi: 10.1097/wnr.0b013e3282f213c5

- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Res.* 1128, 148–156. doi: 10.1016/j.brainres.2006.10.064
- Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., et al. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording and quantifying mismatch negativity, P300 and N400. *Clin. Neurophysiol.* 120, 1883–1908. doi: 10.1016/j.clinph.2009.07.045
- Francis, A. L., Ciocca, V., and Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Percept. Psychophys.* 65, 1029–1044. doi: 10.3758/bf03194832
- Gandour, J. (1983). Tone perception in far eastern languages. *J. Phon.* 11, 149–175.
- Gandour, J. (2006). Brain mapping of Chinese speech prosody. *Handb. East Asian Psycholinguist.* 1, 308–319. doi: 10.1017/cbo9780511550751.030
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dziedzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344–357. doi: 10.1016/j.neuroimage.2004.06.004
- Gandour, J., Wong, D., Hsieh, L., Weinzaepfel, B., Van Lancker, D., and Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *J. Cogn. Neurosci.* 12, 207–222. doi: 10.1162/089892900561841
- Gandour, J., Wong, D., Lowe, M., Dziedzic, M., Sathamnuwong, N., Tong, Y., et al. (2002). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *J. Cogn. Neurosci.* 14, 1076–1087. doi: 10.1162/089892902320474526
- Hallé, P. A., Chang, Y. C., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *J. Phon.* 32, 395–421. doi: 10.1016/s0095-4470(03)00016-0
- Jongman, A., Wang, Y., Moore, C. B., and Sereno, J. A. (2006). “Perception and production of Mandarin Chinese tones,” in *Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)*, eds P. Li, L. Tan, E. Bates and J. L. Tzeng (Cambridge: Cambridge University Press), 209–217.
- Kaan, E., Barkley, C. M., Bao, M., and Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neurosci.* 9:53. doi: 10.1186/1471-2202-9-53
- Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Commun.* 27, 187–207. doi: 10.1016/s0167-6393(98)00085-5
- Klein, D., Zatorre, R. J., Milner, B., and Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage* 13, 646–653. doi: 10.1006/nimg.2000.0738
- Kubovy, M., Cohen, D. J., and Hollier, J. (1999). Feature integration that routinely occurs without focal attention. *Psychon. Bull. Rev.* 6, 183–203. doi: 10.3758/bf03212326
- Li, X., Yang, Y., and Hagoort, P. (2008). Pitch accent and lexical tone processing in Chinese discourse comprehension: an ERP study. *Brain Res.* 1222, 192–200. doi: 10.1016/j.brainres.2008.05.031
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417
- Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., et al. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc. Natl. Acad. Sci. U S A* 103, 19558–19563. doi: 10.1073/pnas.0607065104
- Näätänen, R., and Alho, K. (1997). Mismatch negativity—the measure for central sound representation accuracy. *Audiol. Neurootol.* 2, 341–353. doi: 10.1159/000259255
- Näätänen, R., Gaillard, A. W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol. (Amst)* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9
- Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

- Ren, G. Q., Yang, Y., and Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience* 162, 87–95. doi: 10.1016/j.neuroscience.2009.04.021
- Tsang, Y. K., Jia, S., Huang, J., and Chen, H. C. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: the effects of pitch contour and pitch height. *Neurosci. Lett.* 487, 268–272. doi: 10.1016/j.neulet.2010.10.035
- van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Paper Linguist.* 13, 201–277. doi: 10.1080/08351818009370498
- Wang, X. D., Liu, A. P., Wu, Y. Y., and Wang, P. (2013b). Rapid extraction of lexical tone phonology in Chinese characters: a visual mismatch negativity study. *PLoS One* 8:e56778. doi: 10.1371/journal.pone.0056778
- Wang, X. D., Wang, M., and Chen, L. (2013a). Hemispheric lateralization for early auditory processing of lexical tones: dependence on pitch level and pitch contour. *Neuropsychologia* 51, 2238–2244. doi: 10.1016/j.neuropsychologia.2013.07.015
- Warrier, C. M., and Zatorre, R. J. (2004). Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain* 127, 1616–1625. doi: 10.1093/brain/awh183
- Wong, P. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Res. Bull.* 59, 83–95. doi: 10.1016/s0361-9230(02)00860-2
- Xi, J., Zhang, L., Shu, H., Zhang, Y., and Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience* 170, 223–231. doi: 10.1016/j.neuroscience.2010.06.077
- Xu, Y., Gandour, J. T., and Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120, 1063–1074. doi: 10.1121/1.2213572
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/s1364-6613(00)01816-7
- Zhang, L., Xi, J., Wu, H., Shu, H., and Li, P. (2012a). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *Neuroreport* 23, 35–39. doi: 10.1097/WNR.0b013e32834e4842
- Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., and Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS One* 6:e20963. doi: 10.1371/journal.pone.0020963
- Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., et al. (2012b). Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *J. Child Psychol. Psychiatry* 53, 874–882. doi: 10.1111/j.1469-7610.2012.02528.x
- Zhao, L. (2010). *Experimental Course of ERPs*. Chinese: Southeast University Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 June 2014; accepted: 29 August 2014; published online: 16 September 2014.

Citation: Yu K, Wang R, Li L and Li P (2014) Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity. *Front. Hum. Neurosci.* 8:729. doi: 10.3389/fnhum.2014.00729

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Yu, Wang, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Anterior insular cortex activity to emotional salience of voices in a passive oddball paradigm

Chenyi Chen¹, Yu-Hsuan Lee¹ and Yawei Cheng^{1,2,3*}

¹ Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan

² Department of Rehabilitation, National Yang-Ming University, Yilan, Taiwan

³ Department of Education and Research, Taipei City Hospital, Taipei, Taiwan

Edited by:

Sonja A. Kotz, Max Planck Institute
Leipzig, Germany

Reviewed by:

Sascha Frühholz, University of
Geneva, Switzerland

Alessandro Tavano, University of
Leipzig, Germany

*Correspondence:

Yawei Cheng, Institute of
Neuroscience, National Yang-Ming
University, 155, Sec. 2, St. Linong,
Dist. Beitou, Taipei 112, Taiwan
e-mail: ywcheng2@ym.edu.tw

The human voice, which has a pivotal role in communication, is processed in specialized brain regions. Although a general consensus holds that the anterior insular cortex (AIC) plays a critical role in negative emotional experience, previous studies have not observed AIC activation in response to hearing disgust in voices. We used magnetoencephalography to measure the magnetic counterparts of mismatch negativity (MMNm) and P3a (P3am) in healthy adults while the emotionally meaningless syllables *dada*, spoken as neutral, happy, or disgusted prosodies, along with acoustically matched simple and complex tones, were presented in a passive oddball paradigm. The results revealed that disgusted relative to happy syllables elicited stronger MMNm-related cortical activities in the right AIC and precentral gyrus along with the left posterior insular cortex, supramarginal cortex, transverse temporal cortex, and upper bank of superior temporal cortex. The AIC activity specific to disgusted syllables (corrected $p < 0.05$) was associated with the hit rate of the emotional categorization task. These findings may clarify the neural correlates of emotional MMNm and lend support to the role of AIC in the processing of emotional salience already at the preattentive level.

Keywords: mismatch negativity (MMN), magnetoencephalography (MEG), anterior insular cortex (AIC), emotional salience

INTRODUCTION

Mismatch negativity (MMN) has recently been used as an index of the salience of emotional voice processing (Schirmer et al., 2005; Cheng et al., 2012; Fan et al., 2013; Hung et al., 2013; Fan and Cheng, 2014; Hung and Cheng, 2014). MMN reflects the early saliency detection of auditory stimuli regarding stimulus discrimination based on the perceptual processes of physical features (Pulvermüller and Shtyrov, 2006; Thönnessen et al., 2010). Considering that the anterior insular cortex (AIC) plays a critical role in negative emotional experience (Craig, 2002, 2009), particularly in perceiving disgust, and magnetoencephalography (MEG) could complement the spatiotemporal dynamics in a passive auditory oddball paradigm, we proposed the AIC activation with respect to emotional MMN.

The AIC is a polysensory cortex involved in the awareness of bodily sensations and subjective feelings (Craig, 2002, 2009). Perceiving a disgusting odor, disgusted faces, and imagining feeling disgust have consistently activated the AIC (e.g., Phillips et al., 1997, 2004; Adolphs et al., 2003; Krolak-Salmon et al., 2003; Wicker et al., 2003; Jabbi et al., 2008). Menon and Uddin (2010) suggested that the AIC is a key region of the emotional salience network that integrates external stimuli with internal states to guide behaviors. However, previous studies have failed to identify the AIC activation associated with disgusted vocal expressions (Phillips et al., 1998).

MEG enables non-invasive measurements of neural activity with sufficient spatial resolution and excellent temporal

resolution. MMN/its magnetic equivalent (MMNm), and P3a/P3am, can be elicited using a passive auditory oddball paradigm in which participants engage in a task and must ignore the stimuli that are presented in a random series, with one stimulus (standard) occurring more frequently than the other stimuli (deviant). P3a/P3am is associated with involuntary attention switches for sound changes (Alho et al., 1998). As a preattentive change detection index, MMN/MMNm can reflect N-methyl-D-aspartate receptor function (Näätänen et al., 2011), which mediates sensory memory formation and emotional reactivity in various neuropsychiatric disorders (Campeau et al., 1992; Barkus et al., 2010). MMNm elicited by emotional (happy and angry) deviants in an oddball paradigm can reflect early stimulus processing of emotional prosodies (Thönnessen et al., 2010). Recent studies have indicated that, in addition to being used as an index of the acoustic features of sounds, such as frequency, duration, and phonetic contents (e.g., Ylinen et al., 2006; Horvath et al., 2008), MMN can also be used as an index of the salience of emotional voices (Cheng et al., 2012; Fan et al., 2013; Hung et al., 2013; Fan and Cheng, 2014; Hung and Cheng, 2014).

Various manners have been used for the acoustic control of emotional voices. For example, scrambling voices enables the amplitude envelope to be preserved (Belin et al., 2000). In one study, simple tones synthesized from the strongest formant of the vowel were used as the control stimuli in the mismatch paradigm (Čeponienė et al., 2003). In another studies, physically identical

stimuli were presented as both standards and deviants (Schirmer et al., 2005, 2008). Because no single acoustic parameter can fully explain strong neural responses to emotional prosodies (Wiethoff et al., 2008), the present study, using the same theorems as Belin et al. (2000) did, involved employing two stringent sets of acoustic control stimuli, simple tones and complex tones, to control the temporal envelope and core spectral elements of emotional voices [spectral centroid (fn) and fundamental frequency (f0)], respectively.

To elucidate the neural correlates underpinning the emotional salience of voices, we measured MMNm and P3am in a passive auditory oddball paradigm while presenting the neutrally, happily, and disgustedly spoken syllables *dada* to young adults. We hypothesized that, if the AIC is involved in the preattentive processing of emotional salience, then AIC activation would be observed in the source distribution of MMNm, in accordance with the auditory cortices and early response latencies. If AIC activation is specific to voices, MMNm in response to acoustic attributes, i.e., simple and complex tone deviants, would not elicit the AIC activation. If neurophysiologic changes can guide behaviors (Menon and Uddin, 2010), then people exhibiting stronger AIC activation in response to hearing emotional syllables are expected to perform more favorably in emotional recognition. Furthermore, because men and women might engage in dissimilar neural processing of emotional stimuli (Hamann and Canli, 2004), the gender factor was introduced into the analyses.

MATERIALS AND METHODS

PARTICIPANTS

Twenty healthy participants (10 men), aged 18–30 years (mean \pm SD: 22 ± 1.9), underwent MEG recording and structural MRI scanning after providing written informed consent. The study was approved by the ethics committee in National Yang-Ming University and conducted in accordance with the Declaration of Helsinki. One person was excluded from data analysis because of motion artifacts. All participants were right-handed without hearing or visual impairments. They had no neurological and psychiatric disorders. Participants received monetary compensation for their participation.

AUDITORY STIMULI

The stimulus material consisted of three categories: emotional syllables, simple tones, and complex tones. For the emotional syllables, a young female speaker produced the syllables *dada* with two sets of emotional (happy and disgusted) prosodies and one set of neutral prosodies. Within each type of emotional or neutral prosodies, the speaker produced the *dada* syllables more than ten times to enable validation. Sound Forge 9.0 and Cool Edit Pro 2.0 were used to edit the syllables so that they were equally long (550 ms) and loud (max: 62 dB, mean: 59 dB).

Each syllable set was rated for emotionality on a 5-point Likert-scale by a total of 120 listeners (60 men). For the disgusted set, listeners classified each stimulus from *extremely disgusted* to *not disgusted at all*. For the happy set, listeners classified from *extremely happy* to *not happy at all* and for the neutral set, listeners classified from *extremely emotional* to *not emotional at all*. Emotional syllables that were consistently identified as the

extremely disgusted and happy (i.e., the highest ratings) as well as the most emotionless (i.e., the lowest rating) were used as the stimuli. The Likert-scale (mean \pm SD) of happy, disgusted, and neutral syllables were 4.34 ± 0.65 , 4.04 ± 0.91 , and 2.47 ± 0.87 , respectively.

Although firmly controlling the spectral power distribution may result in the loss of temporal flow associated with formant contents in voices (Belin et al., 2000), the synthesizing the temporal envelope and the core spectral elements of voices should enable the maximal control of the spectral and temporal features of vocal and corresponding non-vocal sounds (Schirmer et al., 2007; Remedios et al., 2009). In order to create a set of stimuli that retain acoustical correspondence with the emotional syllables, we synthesized simple and complex tones by using Praat (Boersma, 2001) and MATLAB (The MathWorks, Inc., Natick, MA, USA). Using a sine waveform, we extracted the fundamental frequencies (f0) and the spectral centroid (fn) of each original syllable to produce complex and simple tones, respectively (Supplementary Figures S1, S2). The lower end of the spectrogram at each time point determined the fundamental frequency (f0). For the complex tones, the f0 over time was extracted to preserve the pitch contour. For the simple tones, the spectral centroid (fn), indicating the center of mass of the spectrum, was extracted to reflect the brightness of sounds. The original syllable envelope then multiplied the extracted frequencies. Hence, to control temporal features, three categories (emotional syllables, complex tones, and simple tones) were assigned to have identical temporal envelopes. To control spectral features, complex tones retained the f0 whereas simple tones retained the fn of emotional syllables. The length (550 ms) and loudness (max: 62 dB; mean: 59 dB) of all stimuli were controlled.

PROCEDURES

During MEG recording, participants lay in a magnetically shielded chamber and watched a silent movie with subtitles while the task-irrelevant vocally spoken or synthesized stimuli were presented. To ensure that the auditory stimuli were sufficiently irrelevant, participants attentively watched the movie and answered questions regarding the movie content after data recording.

Three sessions (emotional syllables, complex tones, and simple tones) were conducted. The session order was pseudorandomized among participants. In the emotional session, neutral syllables set as the standard (S), and happy and disgusted syllables designed as two isometric deviants (D1, D2) followed the oddball paradigm. During the complex and simple sessions, we applied an identical oddball paradigm for the corresponding synthesized tones so that relative acoustic features among S, D1, and D2 were controlled across all three categories. Each session consisted of 800 standards, 100 D1s, and 100 D2s. A minimum of two standards was presented between any two deviants. The successive deviants were always diverse. The stimulus onset asynchrony was 1200 ms.

After MEG recording, participants performed a forced-choice emotional categorization task. While listening to the forty-five stimuli, including five Ss, five D1s, and five D2s of each stimulus category, participants identified each emotional characteristics as one of three types (emotionless, happy, or disgusted) in a

self-paced manner. The chance level was 33.33% based on three alternatives.

APPARATUS AND RECORDINGS

The data were recorded by using a 157-channel axial gradiometer whole-head MEG system (Kanazawa Institute of Technology, Kanazawa, Japan). Prior to data acquisition, the locations of five head position indicator coils attached to the scalp and several additional scalp surface points were recorded with respect to fiducial landmarks (nasion and two preauricular points) by using a 3-D digitizer, which digitized each participant's head shape and localized the position of the participant's head inside the MEG helmet. Data were collected at a sampling frequency of 1 kHz. Participants kept their heads steady during MEG recording. The head-shape and head-position indicator locations were digitized at the onset of recording and were later used to coregister the MEG coordinate system with the structural MRI of each participant. Structural MR images were acquired on a 3 T Siemens Magnetom Trio-Tim scanner using a 3D MPRAGE sequence (TR/TE = 2530/3.5 ms, FOV = 256 mm, flip angle = 7°, matrix = 256 × 256, 176 slices/slab, slice thickness = 1 mm, no gap).

MEG PREPROCESSING AND ANALYSIS

In offline processing the MEG data, we applied a low-pass filter at 20 Hz (Luck, 2005) and reduced the noise using the algorithm of time-shifted principle component analysis (de Cheveigné and Simon, 2007; Hsu et al., 2011). The MEG data were then epoched for each trial type by time locking the stimulus onsets at 100-ms prestimulus intervals and 700-ms poststimulus intervals. Epochs with a signal range exceeding 1.5 fT at any channel were excluded from the averaging and subsequent statistical analyses, in which the deviant-stimulus averages were calculated based on at least 90 trials per participants. The amplitudes of averaged MEG response waveforms were measured with respect to a 100-ms prestimulus baseline.

EVENT-RELATED FIELDS (ERF)

For amplitude and latency analyses, we used the Isofield Contour Map to identify the channels with the strongest signal in the direction. Because head position variation might unequally contribute to the differential activity observed at individual sensors, we created a composite map by grand-averaging nine conditions, where three D1s, three D2s and three Ss of three categories were pooled together. For the ERF difference, the difference maps [disgusted MMNm (D2-S); happy MMNm (D1-S)] were averaged for each category. Based on the composite maps of each ERF component, we selected the four clusters with the strongest signal in the direction. The amplitudes of the sensory ERF peaks (N1m and P2m), MMNm and P3am were measured as an average within a 60-ms window centered at each participant's individual peak latencies. P3am was defined as the component immediately following MMNm, peaking at 300–500 ms. Two-tailed *t* tests were used to determine the statistical presence (difference from 0 fT) of ERF peaks related to the stimuli.

Statistical analysis involved Three-Way mixed ANOVAs with two within-subject factors: category (emotional syllables,

complex tones, or simple tones) and stimulus [neutral (S), happy (D1), or disgusted (D2)] and one between-subject factor: gender (males vs. females). The dependent variables were the amplitudes and latencies of each component. Bonferroni test was conducted only when preceded by significant effects.

MEG SOURCE ANALYSIS

The structural MR images were processed using FreeSurfer (CorTechs Labs, La Jolla, CA and MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA) to create a cortical reconstruction of each brain. Minimum-norm estimates (MNEs) (Hämäläinen and Ilmoniemi, 1994) were computed from combined anatomical MRI and MEG data by using the MNE toolbox (MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA). For inverse computations, the cortical surface was decimated to 5000–10,000 vertices per hemisphere. We used the boundary-element model method to compute the forward solution, which was an estimate of the magnetic field at each MEG sensor resulting from the activity at each of the vertices. The forward solution was then employed to create the inverse solution, which enabled identifying the spatiotemporal distribution of any activity over sources, that most accurately account for each participant's average MEG data. The noise covariance matrix was estimated according to the pre-stimulus baselines of the individual trials. Only the components of activation that were in a direction normal to the cortical surface were retained in the minimum-norm solution. The MNE results were then converted into dynamic statistical parameter maps (dSPM), which measured the noise-normalized activation at each source and enabled several standard minimum-norm calculations inaccuracies to be avoided (Dale et al., 2000).

To test whether the evoked response significantly differed between conditions, the problem of multiple comparisons was addressed by conducting a cluster-level permutation test across space. For each cortical location within each region of interest (ROI), a paired-samples *t* value was computed for testing the deviant-standard contrast or the contrast between two deviants ($p = 0.05$). We then selected all of the samples for which this *t* value exceeded an a priori threshold (uncorrected $p < 0.05$). Finally, the selected samples were clustered according to spatial adjacency. By clustering neighboring cortical locations that exhibited the same effect, we addressed the multiple comparisons problem while considering the dependency of the data. Cortical dipoles were considered to be neighbors if the distance between them was less than 12 mm. A sample was included in the cluster only when there were at least two neighboring samples in space.

REGIONS OF INTERESTS (ROI) AND SOURCE-SPECIFIC TIME-COURSE EXTRACTION

The cortical surface of each participant was normalized onto a standard brain supported by FreeSurfer, and the dSPM solutions of all participants were subsequently averaged so that they could be used in the defined regions and time windows of interest. Considering the fact that the trial number contributed considerably to the inverse source estimation, we selected only the standard exactly preceding the deviant to estimate cortical activity

The dSPM solutions estimated for the standards, which immediately preceded D1 and D2, were then averaged to represent the cortical activity for standard sound processing. To further qualitatively clarify the underlying neural correlates of MMNm and P3am, a functional map, using inclusive masking to display significant dSPM activation for either deviant (D1 or D2), was used to select the ROIs for all of the stimulus categories. Specifically, the grand averaged functional maps evoked by D1 and D2 were overlaid onto a common reconstructed cortical sphere, respectively. The ROIs were drawn along the border of functional maps as well as the anatomical criterion where the vertices were optimally parceled using the gyral-sulcal patterns (Fischl et al., 1999; Sereno et al., 1999; Leonard et al., 2010). The ROIs for D1 and D2 were then combined to form an inclusive mask displaying significant dSPM activations for either deviant (D1 or D2). Then, the dSPM time courses were extracted from the predetermined ROIs after their amplitudes were measured and calculated as described previously (Figure 1).

RESULTS

BEHAVIORAL PERFORMANCE

Table 1 shows the performance on the emotional categorization task. A Three-Way mixed ANOVA targeting categories (emotional syllables, complex tones, or simple tones) and stimulus (neutral, happy, or disgusted) as the within-subject variables and gender (male vs. female) as the between-subject variable was computed in terms of the hit rate. The category effect [$F_{(2, 32)} = 4.90$, $p = 0.01$] and the interaction between category and stimulus [$F_{(4, 64)} = 2.80$, $p = 0.03$] were significant (Figure 2). No significance was observed regarding gender and gender-related interaction. The emotional syllables exhibited more favorable

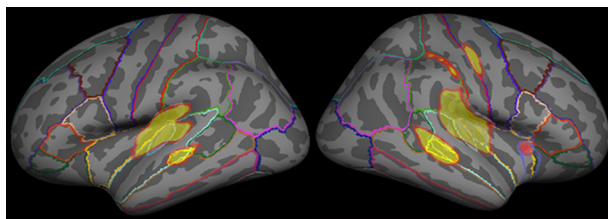


FIGURE 1 | The grand averaged functional map evoked by emotional syllables. The functional map of dSPM solutions estimated for each category (emotional syllables, complex, and simple tones) was superimposed onto a reconstructed anatomical criterion where the vertices were optimally parceled out using the gyral-sulcal patterns. We overlaid the grand averaged functional map in response to emotional syllables onto the common reconstructed cortical sphere as an example.

Table 1 | Behavioral performance.

Emotional categorization task	Mean \pm SD		
	Emotional syllables	Complex tones	Simple tones
Accuracies (%)	61.5 \pm 28.8	48.5 \pm 20.4	44.8 \pm 15.1
Response times (ms)	798 \pm 422	791 \pm 297	731 \pm 252

performance than did the complex ($p = 0.031$) and simple ($p = 0.005$) tones. *Post-hoc* comparisons showed that the neural relative to emotional-derived tones exerted more favorable performance in the complex [Neutral- > Happy-derived tones: $t_{(17)} = 1.9$, Cohen's $d = 0.70$, one-tailed $p = 0.035$; Neutral- > Disgusted-derived tones: $t_{(17)} = 2.5$, Cohen's $d = 0.92$, one-tailed $p = 0.01$] and simple tones [$t_{(17)} = 1.8$, Cohen's $d = 0.67$, one-tailed $p = 0.04$; $t_{(17)} = 1.8$, Cohen's $d = 0.67$, one-tailed $p = 0.04$], but this pattern was not observed in the emotional syllables ($p = 0.11$; $p = 0.94$). Only the emotional syllables exerted above-chance hit rates (>33.33%) in all emotions, rather than the complex and simple tones, indicating emotional neutrality of acoustic controls.

SENSORY ERF

Each stimulus type of each category reliably elicited an N1-P2 complex, which is typically obtained in adults during fast stimulus presentation (Supplementary Table S1 and Figure S3) (Näätänen and Picton, 1987; Pantev et al., 1988; Tremblay et al., 2001; Shahin et al., 2003; Ross and Tremblay, 2009).

Statistical analyses for each identified cluster on the Isofield Contour Map revealed that N1m had the category effect for the cluster over the right anterior region [$F_{(2, 36)} = 5.33$, $p = 0.009$] and P2m had the category effect for the clusters over the left and right posterior regions [$F_{(2, 36)} = 15.03$, $p < 0.001$; $F_{(2, 36)} = 13.65$, $p < 0.001$]. *Post-hoc* analyses indicated that the emotional syllables elicited stronger N1m than did the complex tones ($p = 0.005$) and simple tones ($p = 0.02$) for the cluster over the right anterior region. For those over the left and right posterior regions, the simple tones elicited stronger P2m amplitudes than did the emotional syllables (left: $p = 0.03$; right: $p = 0.02$) and complex tones ($p = 0.01$; $p < 0.001$). There was no significance for gender and gender-related interaction.

MMNm

All deviant stimuli of each category elicited MMNm significantly. Statistical analyses for each identified cluster on the Isofield Contour Map revealed that there were significant interactions between category and stimulus over the left [$F_{(2, 36)} = 4.73$,

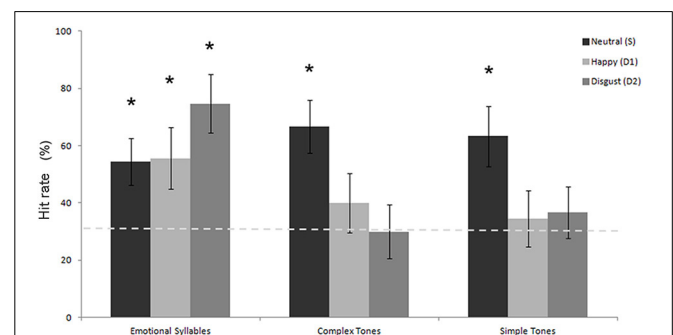


FIGURE 2 | Hit rate in the emotional categorization task. The emotional syllables exhibited more favorable performance than did the complex and simple tones. Only the emotional syllables attained above-chance hit rates for each emotion. The asterisk (* $p < 0.05$) indicates that the hit rate is statistically higher than the chance level (dashed line).

$p = 0.015$] and right posterior [$F_{(2, 36)} = 5.05, p = 0.012$] clusters. In the left posterior cluster, *post-hoc* analyses showed that the stimulus effect where disgusted (D2) relative to happy (D1) MMNm was larger in amplitudes was present in the emotional syllables ($p = 0.004$), but none was detected in the simple tones ($p = 0.27$) and complex tones ($p = 0.41$). In the right posterior cluster, the stimulus effect was found in the emotional syllables ($p < 0.001$) and simple tones ($p = 0.005$). Gender and gender-related interaction were not significant ($p > 0.05$).

MMNm-RELATED CORTICAL ACTIVITIES

Table 2 lists the peak latencies used for analyzing source-specific amplitudes for each ROI. Statistical analyses, using a Three-Way mixed ANOVA targeting category (emotional syllables, complex tones, or simple tones) and stimulus [neutral (S), happy (D1), or disgusted (D2)] as the within-subject variables and gender (males vs. females) as the between-subject variable for each ROI, revealed that the brain regions exhibiting an interaction between the category and stimulus were the right AIC [$F_{(4, 68)} = 3.35, p = 0.015$], right precentral gyrus [$F_{(4, 68)} = 5.54, p = 0.001$], left supramarginal cortex [$F_{(4, 68)} = 3.03, p = 0.023$], upper and lower bank of superior temporal sulcus (uSTS and lSTS) [$F_{(4, 68)} = 2.78, p = 0.03$; $F_{(4, 68)} = 3.41, p = 0.03$] together with the left posterior insular cortex (PIC) [$F_{(4, 68)} = 5.09, p = 0.006$], left and right transverse temporal cortex [$F_{(4, 68)} = 4.71, p = 0.002$; $F_{(4, 68)} = 3.02, p = 0.02$] (**Figure 3**). Gender and gender-related interaction were non-significant. *Post-hoc* tests indicated that the processing of emotional salience (**Figure 4**), as indicated by stronger cortical activities for disgusted syllables relative to happy syllables (D2 > D1; FDR corrected $p < 0.05$) occurred in the right AIC [$F_{(2, 34)} = 7.84, p = 0.002$] and precentral gyrus [$F_{(2, 34)} = 7.98, p = 0.005$] along with the left PIC [$F_{(2, 34)} = 5.55, p = 0.008$], supramarginal cortex [$F_{(2, 34)} = 6.71, p = 0.004$], transverse temporal cortex [$F_{(2, 34)} = 8.95, p = 0.001$], and uSTS [$F_{(2, 34)} = 8.49, p = 0.001$].

In addition, the right AIC activities, which surpassed the dSPM criterion, specifically responded to the disgusted syllables (**Figure 5**). The correlation analysis revealed that the MMNm-related AIC activities were associated with the hit rates for emotional syllables in the emotional categorization task [$r_{(18)} = 0.49, p = 0.036$]. Participants exhibiting larger amplitudes in the right AIC activation triggered by disgusted syllables were likely to perform better in the emotional categorization task.

P3am

Paired *t*-tests used to determine the statistical presence (difference from 0 fT/cm) indicated that all deviants from each category elicited P3am, temporally following MMNm (Supplementary Table S2). The ANOVA model on P3am amplitudes for each identified cluster on the Isofield Contour Map, did not find any significance in the category, stimulus, gender, and their related interaction (all $p > 0.05$).

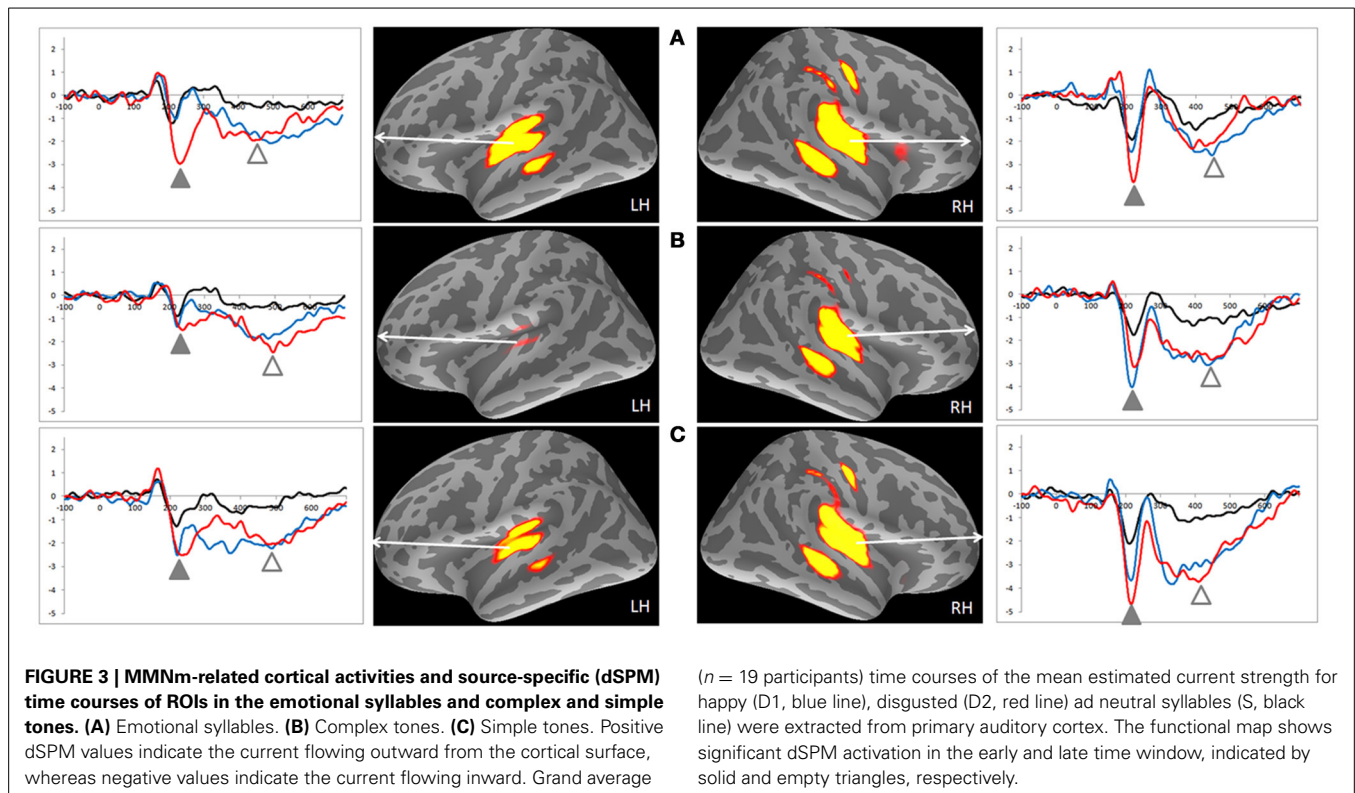
P3am-RELATED CORTICAL ACTIVITIES

P3am and MMNm had similar brain sources and manifested as two contiguous peaks in the dSPM source-specific time course (**Figure 3** and **Table 3**). Statistical analyses on P3am-related cortical activities used a Three-Way mixed ANOVA targeting category (emotional syllables, complex tones, or simple tones) and stimulus [neutral (S), happy (D1), or disgusted (D2)] as the within-subject variables and gender (male vs. female) as the between-subject variable. The brain regions exhibiting the stimulus effect included the right supramarginal cortex [$F_{(2, 34)} = 25.37, p = 0.032$], uSTS [$F_{(2, 34)} = 23.48, p = 0.018$], lSTS [$F_{(2, 34)} = 30.94, p = 0.006$], and posterior superior temporal sulcus (pSTS) [$F_{(2, 34)} = 17.44, p = 0.004$], as well as the left transverse temporal cortex [$F_{(2, 34)} = 16.67, p = 0.002$], supramarginal cortex [$F_{(2, 34)} = 12.00, p = 0.008$], PIC [$F_{(2, 34)} = 16.25, p = 0.006$], uSTS [$F_{(2, 34)} = 14.81, p = 0.003$], and lSTS [$F_{(2, 34)} = 26.42, p = 0.001$]. None of ROIs reached

Table 2 | MMNm-related cortical activities.

Regions of interest (ROI)	L/R	MNI coordinates			Peak latencies (ms; mean \pm SD)			Post-hoc FDR < 0.05		
		X	Y	Z	Emotional	Complex	Simple	Emotional	Complex	Simple
Anterior insular cortex (AIC)	R	30	16	4	240 \pm 24			D2>S, D2>D1		
Posterior insular cortex (PIC)	L	-31	-22	10	241 \pm 22			D2>D1		
Superior temporal sulcus, upper bank (uSTS)	L	-50	-26	2	244 \pm 27	248 \pm 25	248 \pm 27	D2>S, D2>D1	D2>S	D2>S
Supramarginal cortex	L	-42	-24	22	240 \pm 23	245 \pm 20		D2>D1	D2>S	
Transverse temporal cortex	L	-45	-22	7	245 \pm 25	248 \pm 24		D2>D1	D2>S	
Transverse temporal cortex	R	49	-19	4		241 \pm 23	233 \pm 20		D1>S	D2>S
Superior temporal sulcus, upper bank (uSTS)	L	-50	-27	0	244 \pm 26		248 \pm 26	D2>S, D2>D1		D2>S
Precentral gyrus	R	44	-16	36	230 \pm 23		232 \pm 18	D2>D1		D1>S, D2>S
Posterior insular cortex (PIC)	R	34	-16	0			233 \pm 22			D2>S
Superior temporal sulcus, lower bank (lSTS)	R	48	-28	2		240 \pm 23			D1>S	

S, standard (neutral); D1, happy; D2, disgusted.



any interaction between the category and stimulus. Gender and gender-related interaction were non-significant.

DISCUSSION

Although the AIC plays a critical role in the negative experience of emotions (Craig, 2002, 2009; Menon and Uddin, 2010), including disgust (e.g., Phillips et al., 1997, 2004; Adolphs et al., 2003; Krolak-Salmon et al., 2003; Wicker et al., 2003; Jabbi et al., 2008), previous studies have not observed AIC activation in response to hearing disgusted voices. It thus leaves a room for more research to clarify whether AIC activation is specific to disgust or, alternatively, reflects general aversive arousal in response to negative emotions. In contrast to the predicted correlation between AIC activation and disgust recognition, we determined that the AIC activation predicted the performance of emotion recognition in general within the emotional category. This may be partially attributed to the functional role of the AIC in salience processing (Seeley et al., 2007; Sridharan et al., 2008; Menon and Uddin, 2010; Legrain et al., 2011). One MEG study on disgusted faces reported that early insular activation occurs at approximately 200 ms after the stimulus onset of emotionally arousing stimuli, regardless of valence, whereas the later insular response (350 ms) differentiates disgusted from happy facial expressions (Chen et al., 2009). Accordingly, AIC activation occurred in response to disgusted voices at approximately 250 ms in our study. The AIC is a brain region underpinning error awareness and saliency detection (Sterzer and Kleinschmidt, 2010; Harsay et al., 2012). This passive oddball study required no target detection. The AIC activation, which surpassed the dSPM criterion, specifically responded

to disgusted syllables rather than happy syllables. Participants exhibiting stronger AIC activities were likely to have higher hit rates in the emotional categorization task (please see Figure 5). In addition, disgusted relative to happy syllables exhibited stronger MMNm-related cortical activities, lending support for the notion that disgusted relative to happy voices might be more acoustically salient (Banse and Scherer, 1996; Simon-Thomas et al., 2009; Sauter et al., 2010).

Using MEG in a passive auditory oddball paradigm, we demonstrated the involvement of AIC in the preattentive perception of disgusted voices. MMNm-related AIC activation was specific to disgusted syllables, but not happy syllables. In addition, acoustically matched simple and complex tones did not activate the AIC in the same manner. Participants who exhibited stronger MMNm-related AIC activations were more prone to obtaining higher hit rates in the emotional categorization task. The involvement of AIC in emotional MMNm appears to be consistent between genders.

The MMNm response was sensitive to the positive and negative valence of emotional voices, as indicated by stronger amplitudes elicited by disgusted syllables than by happy syllables. Particularly, in the left hemisphere, the emotional salience processing was specific to voices rather than their acoustic attributes (see Figure 4). This should not be surprising because affective discrimination beyond acoustical distinction emerges early in the neonatal period (Cheng et al., 2012). Hearing angry and fearful syllables relative to happy syllables elicited stronger MMNm (Schirmer et al., 2005; Fan et al., 2013; Hung et al., 2013; Fan and Cheng, 2014; Hung and Cheng, 2014). From an evolutionary

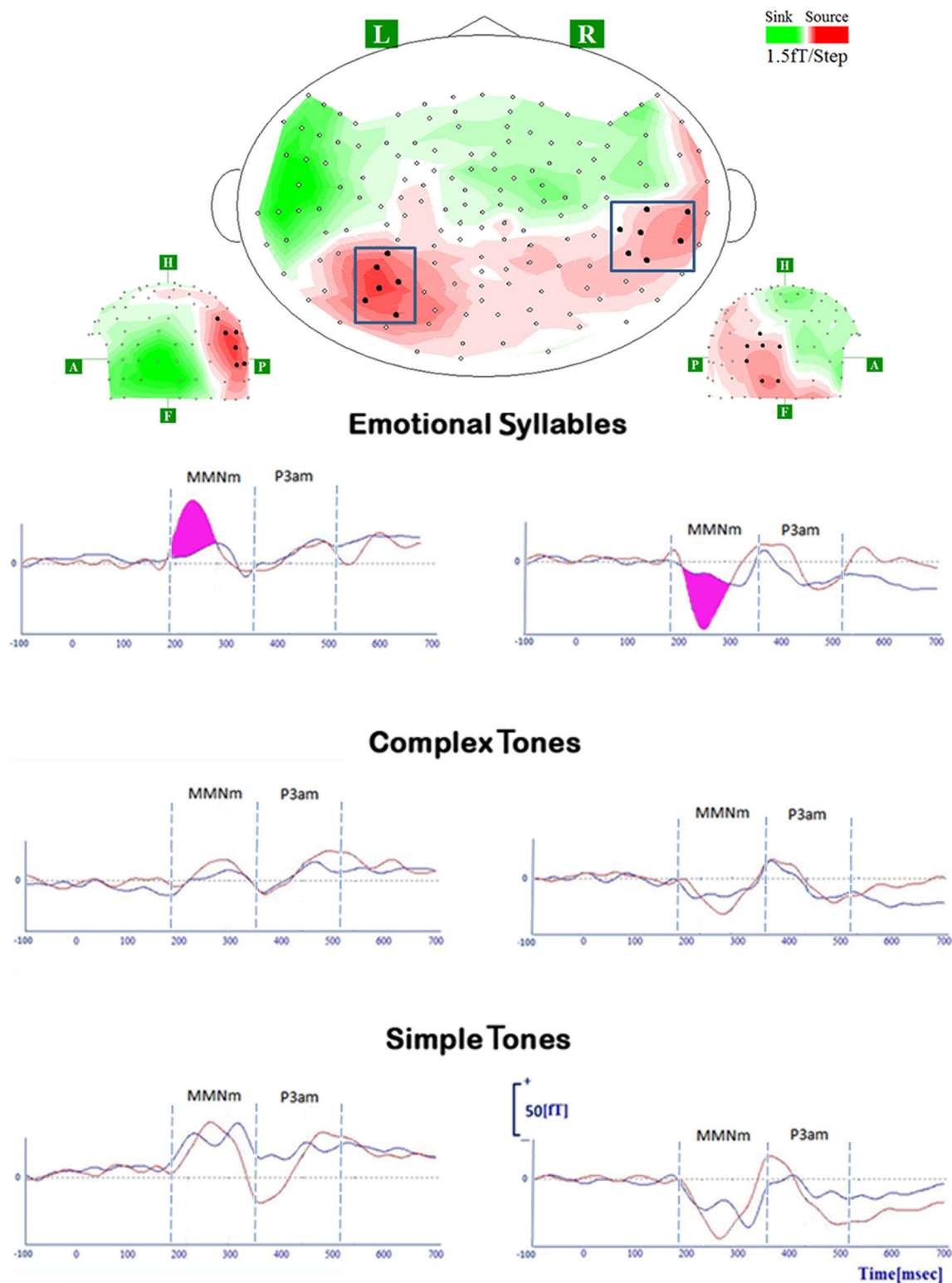


FIGURE 4 | MMNm and P3am in response to deviant stimuli of emotional syllables, complex tones, and simple tones. All deviant stimuli of each category (emotional syllables, complex, and simple tones) consistently elicit MMNm and P3am [red line: disgusted MMNm (D2-S); blue line: happy MMNm

(D1-S)]. For the sake of argument, the grand-average whole-head topography was derived from emotional syllables at the peak latency of MMNm (265 ms). The magenta area indexes the emotional salience processing, as shown by significant differences between disgusted MMNm and happy MMNm.

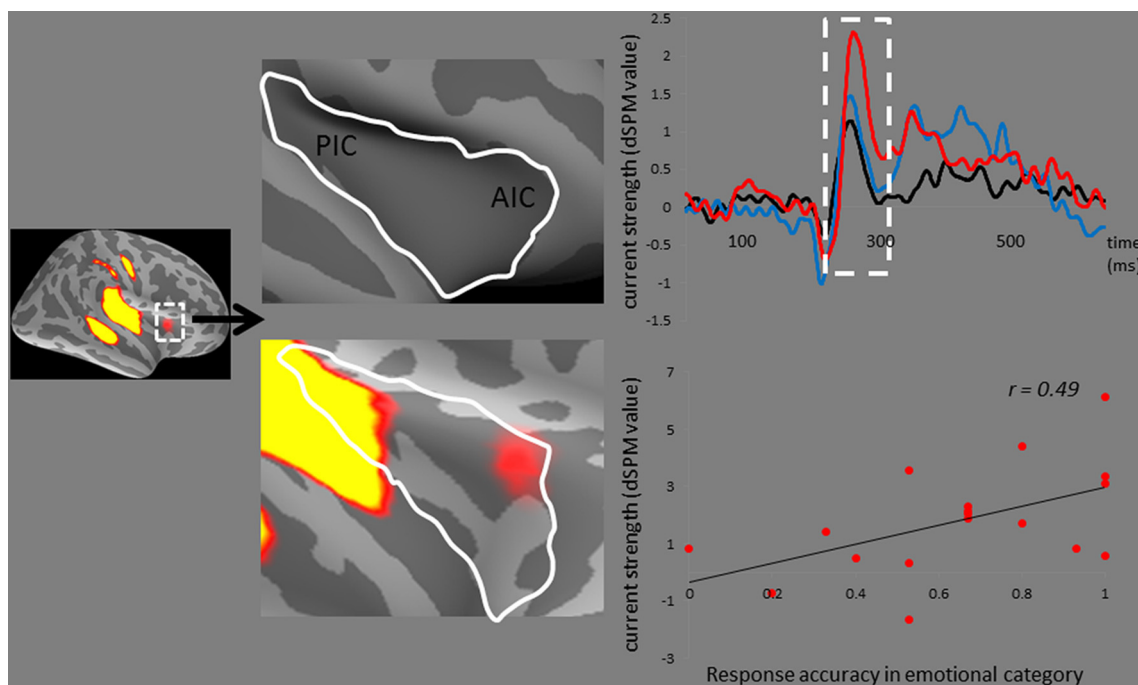


FIGURE 5 | The anterior insular cortex activity in response to disgusted syllables. Disgusted deviants exclusively elicited the activation in the AIC, as indicated by MMNm-related cortical activities. Grand average ($n = 19$ participants) time courses of the mean estimated current strength were

extracted from the right AIC (red line: disgust, D2; blue line: happy, D1; black line: standard, S). The AIC activation for disgusted deviants and the hit rates in the emotional categorization task were positively correlated [$r_{(18)} = 0.49$, $p = 0.036$].

Table 3 | P3am-related cortical activities.

Regions of interest	L/R	MNI coordinates			Peak latencies (ms; mean \pm SD)			Post-hoc FDR < 0.05		
		X	Y	Z	Emotional	Complex	Simple	Emotional	Complex	Simple
Supramarginal cortex	R	46	-24	20	383 \pm 36	389 \pm 46	379 \pm 35	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Superior temporal sulcus, upper bank (uSTS)	R	52	-29	9	363 \pm 34	380 \pm 43	370 \pm 38	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Superior temporal sulcus, lower bank (lSTS)	R	48	-28	2	386 \pm 32	392 \pm 40	380 \pm 37	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Supramarginal cortex	L	-42	-24	22	408 \pm 44	419 \pm 41	409 \pm 52	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Transverse temporal cortex	L	-45	-22	7	415 \pm 43	420 \pm 42	411 \pm 49	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Superior temporal sulcus, upper bank (uSTS)	L	-50	-26	2	405 \pm 46	409 \pm 45	411 \pm 51	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Posterior insular cortex (PIC)	L	-31	-22	10	412 \pm 42	432 \pm 42	412 \pm 42	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Superior temporal sulcus, lower bank (lSTS)	L	-50	-27	0	420 \pm 45	427 \pm 51	414 \pm 48	D1>S, D2>S	D1>S, D2>S	D1>S, D2>S
Posterior superior temporal sulcus (pSTS)	R	43	-38	1	364 \pm 35	366 \pm 36	360 \pm 37	D1>S, D2>S	D1>S, D2>S	D2>S

S, neutral; D1, happy; D2, disgusted.

perspective, disgust, an aversive emotion, exhibits a negativity bias that elicits stronger responses than neutral events do (Lange, 1922; Huang and Luo, 2006).

MMNm-related AIC activation may reflect emotional salience at the preattentive level. The significance of this finding lies in

the lack of similar findings in the auditory modality by previous studies, despite the widespread evidence of the involvement of AIC in the experience of negative emotions. Particularly, attentively hearing disgusted voices did not activate the AIC (Phillips et al., 1998), indicating that AIC may be involved in the

preattentive processing of disgusted voices. Theoretically, the passive auditory oddball paradigm should be the optimal approach for probing the preattentive processing of emotional voices because MMN can indicate the neural activity in a comatose or deep-sleeping brain (Kotchoubey et al., 2002). Although a silent movie presentation is unable to guarantee the lack of awareness to auditory stimuli, limited attention resources can indeed modulate the neurophysiological processing of emotional stimuli (Pessoa and Adolphs, 2010). We do not assert that the preattentive processing of emotional salience of voices was only dominated by the AIC. Through the AIC, the cortical-subcortical interactions for coordinating the function of cortical networks might be attributed to the neural mechanism underpinning the evaluation of the biological significance of affective voices.

The PIC exhibited stronger MMNm-related cortical activities for disgusted syllables relative to happy syllables, possibly involved in the representation of emotional salience. The PIC that has been functionally identified as the portion of the extended auditory cortex responded preferentially to vocal communication sounds (Remedios et al., 2009). The salient sensory information would reach the multimodal cortical areas, such as the PIC, directly from the thalamus, bypassing primary sensory cortices. This direct thalamocortical transmission is parallel to the modality-specific processing of stimulus attributes via the transmission from the thalamus to the relevant primary sensory cortices (Liang et al., 2013). In the present study, failing to observe any activation in the thalamus and anterior cingulate cortex in every condition could be attributed to the stringent statistic dSPM criterion as well as the absence of the target detection in a passive oddball task. The insular cortex, including AIC and PIC, can monitor the salience (appetitive and aversive) and integrate with the stimulus effect on the state of the body (Deen et al., 2011).

In addition, the left transverse temporal cortex, a part of primary auditory cortex, was sensitive to the processing of emotional salience [disgust (D2) > happy (D1)]. This finding supports that vocal emotional expression might be processed beyond the right hemisphere, being anchored within sensory, cognitive, and emotional processing systems at an early auditory discrimination stage (Schirmer and Kotz, 2006). The activation of the precentral gyrus observed across all three categories may reflect general attention and memory enhancement during information processing (Chen et al., 2010).

Importantly, the present study identified several future areas of inquiry. First, familiarity might potentially confound the effect of affective modulation observed here. Simple and complex tones are less familiar than emotional voices, and it might be impossible to categorize synthesized tones in relation to natural speech sounds. Second, using a pseudoword, such as *dada*, as an example might limit the generalizability for emotion representation. By using non-linguistic emotional vocalizations (Fecteau et al., 2007), additional studies are needed to verify whether the passive oddball paradigm is optimal for detecting emotional salience. Third, based on three alternatives, we defined the hit rate as the number of hits divided by the total number of trials at each stimulus category. This study did not control false alarm rates with the traditional approaches within the framework of Signal Detection Theory [SDT: $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$]. The

performance of acoustic controls showed the skewed distribution, in which participants tended to classify the emotional-derived tones as neutral. Accordingly, a higher hit rate for neutral-derived tones than happy-/disgusted-derived tones might potentially violate the assumption that acoustic controls are emotionless. On the other hand, the sum of the hit rate for emotional syllables in the emotional categorization task, which prevents false alarms, exhibited a better prediction to the AIC activation than did the hit rate specific for disgust syllables. The hit rate of 75% for disgusted syllables and 55% for happy syllables corroborated existing findings in the identification of prototypical disgust and pleasure vocal burst (Simon-Thomas et al., 2009). However, only 55% being the hit rate for happy syllables has to be interpreted with caution. Future research should include the pre-evaluation of voices not only on the target emotion but also on other valence scales.

This MEG study clearly demonstrated the right AIC activation in response to disgusted deviance in a passive auditory oddball paradigm. The MMNm-related AIC activity was associated with the emotional categorization performance. The findings may clarify the neural correlates of emotional MMN and support that the AIC is involved in the processing of emotional salience at the preattentive level.

ACKNOWLEDGMENTS

The study was funded by the Ministry of Science and Technology (MOST 103-2401-H-010-003-MY3), National Yang-Ming University Hospital (RD2014-003), Health Department of Taipei City Government (10301-62-009), and Ministry of Education (Aim for the Top University Plan). None of the authors have any conflicts of interest.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2014.00743/abstract>

REFERENCES

- Adolphs, R., Tranel, D., and Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain Cogn.* 52, 61–69. doi: 10.1016/S0278-2626(03)00009-5
- Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jaaskelainen, I. P., et al. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology* 35, 211–224. doi: 10.1111/1469-8986.3520211
- Banase, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. doi: 10.1037/0022-3514.70.3.614
- Barkus, C., McHugh, S. B., Sprengel, R., Seeburg, P. H., Rawlins, J. N., and Bannerman, D. M. (2010). Hippocampal NMDA receptors and anxiety: at the interface between cognition and emotion. *Eur. J. Pharmacol.* 626, 49–56. doi: 10.1016/j.ejphar.2009.10.014
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott. Int.* 5, 341–345. Available online at: <http://www.fon.hum.uva.nl/praat/>
- Campeau, S., Miserendino, M. J., and Davis, M. (1992). Intra-amygdala infusion of the N-methyl-D-aspartate receptor antagonist AP5 blocks acquisition but not expression of fear-potentiated startle to an auditory conditioned stimulus. *Behav. Neurosci.* 106, 569–574. doi: 10.1037/0735-7044.106.3.569
- Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., et al. (2003). Speech-sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5567–5572. doi: 10.1073/pnas.0835631100

- Chen, T. L., Babiloni, C., Ferretti, A., Perrucci, M. G., Romani, G. L., Rossini, P. M., et al. (2010). Effects of somatosensory stimulation and attention on human somatosensory cortex: an fMRI study. *Neuroimage* 53, 181–188. doi: 10.1016/j.neuroimage.2010.06.023
- Chen, Y. H., Dammers, J., Boers, F., Leiber, S., Edgar, J. C., Roberts, T. P., et al. (2009). The temporal dynamics of insula activity to disgust and happy facial expressions: a magnetoencephalography study. *Neuroimage* 47, 1921–1928. doi: 10.1016/j.neuroimage.2009.04.093
- Cheng, Y., Lee, S. Y., Chen, H. Y., Wang, P. Y., and Decety, J. (2012). Voice and emotion processing in the human neonatal brain. *J. Cogn. Neurosci.* 24, 1411–1419. doi: 10.1162/jocn_a_00214
- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894
- Craig, A. D. (2009). How do you feel-now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67. doi: 10.1016/S0896-6273(00)81138-1
- de Cheveigné, A., and Simon, J. Z. (2007). Denoising based on time-shift PCA. *J. Neurosci. Methods* 165, 297–305. doi: 10.1016/j.jneumeth.2007.06.003
- Deen, B., Pitskel, N. B., and Pelphrey, K. A. (2011). Three systems of insular functional connectivity identified with cluster analysis. *Cereb. Cortex* 21, 1498–1506. doi: 10.1093/cercor/bhq186
- Fan, Y. T., and Cheng, Y. (2014). Atypical mismatch negativity in response to emotional voices in people with autism spectrum conditions. *PLoS ONE* 9:e102471. doi: 10.1371/journal.pone.0102471
- Fan, Y. T., Hsu, Y. Y., and Cheng, Y. (2013). Sex matters: n-back modulates emotional mismatch negativity. *Neuroreport* 24, 457–463. doi: 10.1097/WNR.0b013e32836169b9
- Fecteau, S., Belin, P., Joanette, Y., and Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36, 480–487. doi: 10.1016/j.neuroimage.2007.02.043
- Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284. doi: 10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4
- Hämäläinen, M., and Ilmoniemi, R. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42. doi: 10.1007/BF02512476
- Hamann, S., and Canli, T. (2004). Individual differences in emotion processing. *Curr. Opin. Neurobiol.* 14, 233–238. doi: 10.1016/j.conb.2004.03.010
- Harsay, H. A., Spaan, M., Wijnen, J. G., and Ridderinkhof, K. R. (2012). Error awareness and salience processing in the oddball task: shared neural mechanisms. *Front. Hum. Neurosci.* 6:246. doi: 10.3389/fnhum.2012.00246
- Horvath, J., Winkler, I., and Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biol. Psychol.* 79, 139–147. doi: 10.1016/j.biopsycho.2008.04.001
- Hsu, C. H., Lee, C. Y., and Marantz, A. (2011). Effects of visual complexity and sublexical information in the occipitotemporal cortex in the reading of Chinese phonograms: a single-trial analysis with MEG. *Brain Lang.* 117, 1–11. doi: 10.1016/j.bandl.2010.10.002
- Huang, Y. X., and Luo, Y. J. (2006). Temporal course of emotional negativity bias: an ERP study. *Neurosci. Lett.* 398, 91–96. doi: 10.1016/j.neulet.2005.12.074
- Hung, A. Y., Ahveninen, J., and Cheng, Y. (2013). Atypical mismatch negativity to distressful voices associated with conduct disorder symptoms. *J. Child Psychol. Psychiatry* 54, 1016–1027. doi: 10.1111/jcpp.12076
- Hung, A. Y., and Cheng, Y. (2014). Sex differences in preattentive perception of emotional voices and acoustic attributes. *Neuroreport* 25, 464–469. doi: 10.1097/WNR.0000000000000115
- Jabbi, M., Bastiaansen, J., and Keyers, C. (2008). A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. *PLoS ONE* 3:e2939. doi: 10.1371/journal.pone.0002939
- Kotchoubey, B., Lang, S., Bostanov, V., and Birbaumer, N. (2002). Is there a mind? Electrophysiology of unconscious patients. *News Physiol. Sci.* 17, 38–42. Available online at: <http://physiologyonline.physiology.org/content/17/1/38.article-info>
- Krolak-Salmon, P., Henaff, M. A., Isnard, J., Tallon-Baudry, C., Guenot, M., Vighetto, A., et al. (2003). An attention modulated response to disgust in human ventral anterior insula. *Ann. Neurol.* 53, 446–453. doi: 10.1002/ana.10502
- Lange, C. G. (1922). “The emotions: a psychological study,” in *The Emotions*, eds C. G. Lange and W. James (Baltimore, MD: Williams & Wilkins), 33–90.
- Legrain, V., Iannetti, G. D., Plaghki, L., and Mouraux, A. (2011). The pain matrix reloaded: a salience detection system for the body. *Prog. Neurobiol.* 93, 111–124. doi: 10.1016/j.pneurobio.2010.10.005
- Leonard, M. K., Brown, T. T., Travis, K. E., Gharapetian, L., Hagler, D. J. Jr., Dale, A. M., et al. (2010). Spatiotemporal dynamics of bilingual word processing. *Neuroimage* 49, 3286–3294. doi: 10.1016/j.neuroimage.2009.12.009
- Liang, M., Mouraux, A., and Iannetti, G. D. (2013). Bypassing primary sensory cortices—a direct thalamocortical pathway for transmitting salient sensory information. *Cereb. Cortex* 23, 1–11. doi: 10.1093/cercor/bhr363
- Luck, S. J. (2005). *An Introduction To The Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi: 10.1007/s00429-010-0262-0
- Näätänen, R., Kujala, T., Kreegipuu, K., Carlson, S., Escera, C., Baldeweg, T., et al. (2011). The mismatch negativity: an index of cognitive decline in neuropsychiatric and neurological diseases and in ageing. *Brain* 134, 3435–3453. doi: 10.1093/brain/awr064
- Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x
- Pantev, C., Hoke, M., Lehnertz, K., Lutkenhoner, B., Anogianakis, G., and Wittkowski, W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalogr. Clin. Neurophysiol.* 69, 160–170. doi: 10.1016/0013-4694(88)90211-8
- Pessoa, L., and Dolphs, R. (2010). Emotion processing and the amygdala: from a “low road” to “many roads” of evaluating biological significance. *Nat. Rev. Neurosci.* 11, 773–783. doi: 10.1038/nrn2920
- Phillips, M. L., Williams, L. M., Heining, M., Herba, C. M., Russell, T., Andrew, C., et al. (2004). Differential neural responses to overt and covert presentations of facial expressions of fear and disgust. *Neuroimage* 21, 1484–1496. doi: 10.1016/j.neuroimage.2003.12.013
- Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc. Biol. Sci.* 265, 1809–1817. doi: 10.1098/rspb.1998.0506
- Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389, 495–498. doi: 10.1038/39051
- Pulvermüller, F., and Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 79, 49–71. doi: 10.1016/j.pneurobio.2006.04.004
- Remedios, R., Logothetis, N. K., and Kayser, C. (2009). An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *J. Neurosci.* 29, 1034–1045. doi: 10.1523/JNEUROSCI.4089-08.2009
- Ross, B., and Tremblay, K. (2009). Stimulus experience modifies auditory neuro-magnetic responses in young and older listeners. *Hear. Res.* 248, 48–59. doi: 10.1016/j.heares.2008.11.012
- Sauter, D. A., Eisner, F., Calder, A. J., and Scott, S. K. (2010). Perceptual cues in non-verbal vocal expressions of emotions. *Q. J. Exp. Psychol. (Hove)*. 63, 2251–2272. doi: 10.1080/17470211003721642
- Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., and Friederici, A. D. (2008). When vocal processing gets emotional: on the role of social orientation in relevance detection by the human amygdala. *Neuroimage* 40, 1402–1410. doi: 10.1016/j.neuroimage.2008.01.018
- Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009
- Schirmer, A., Simpson, E., and Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Res.* 1176, 103–112. doi: 10.1016/j.brainres.2007.08.008
- Schirmer, A., Striano, T., and Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport* 16, 635–639. doi: 10.1097/00001756-200504250-00024

- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007
- Sereno, M. I., Dale, A. M., Liu, A. K., and Tootell, R. B. H. (1999). A surface-based coordinate system for a canonical cortex. *Neuroimage* 9, 195–207. doi: 10.1016/S1053-8119(96)80254-0
- Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552. Available online at: <http://www.jneurosci.org/content/23/13/5545.short>
- Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., and Abramson, A. (2009). The voice conveys specific emotions: evidence from vocal burst displays. *Emotion* 9, 838–846. doi: 10.1037/a0017810
- Sridharan, D., Levitin, D. J., and Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12569–12574. doi: 10.1073/pnas.0800005105
- Sterzer, P., and Kleinschmidt, A. (2010). Anterior insula activations in perceptual paradigms: often observed but barely understood. *Brain Struct. Funct.* 214, 611–622. doi: 10.1007/s00429-010-0252-2
- Thönnessen, H., Boers, F., Dammers, J., Chen, Y. H., Norra, C., and Mathiak, K. (2010). Early sensory encoding of affective prosody: neuromagnetic tomography of emotional category changes. *Neuroimage* 50, 250–259. doi: 10.1016/j.neuroimage.2009.11.082
- Tremblay, K., Kraus, N., McGee, T., Ponton, C., and Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear.* 22, 79–90. doi: 10.1097/00003446-200104000-00001
- Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664. doi: 10.1016/S0896-6273(03)00679-2
- Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., et al. (2008). Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. *Neuroimage* 39, 885–893. doi: 10.1016/j.neuroimage.2007.09.028
- Ylinen, S., Shestakova, A., Huottilainen, M., Alku, P., and Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain Res.* 1072, 175–185. doi: 10.1016/j.brainres.2005.12.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 March 2014; accepted: 03 September 2014; published online: 22 September 2014.

Citation: Chen C, Lee Y-H and Cheng Y (2014) Anterior insular cortex activity to emotional salience of voices in a passive oddball paradigm. *Front. Hum. Neurosci.* 8:743. doi: 10.3389/fnhum.2014.00743

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Chen, Lee and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Talking hands: tongue motor excitability during observation of hand gestures associated with words

Naeem Komeilipoor^{1,2}, Carmelo Mario Vicario³, Andreas Daffertshofer^{2*} and Paola Cesari¹

¹ Department of Neurological and Movement Sciences, University of Verona, Verona, Italy

² MOVE Research Institute Amsterdam, VU University Amsterdam, Amsterdam, Netherlands

³ School of Psychology, Bangor University, Bangor, UK

Edited by:

Carolyn McGettigan, Royal Holloway University of London, UK

Reviewed by:

Claudia Gianelli, University of Potsdam, Germany

Benjamin Straube, Philipps University, Germany

*Correspondence:

Andreas Daffertshofer, MOVE Research Institute Amsterdam, VU University Amsterdam, Van der Boechorststraat 9, Amsterdam, 1081 BT, Netherlands
e-mail: a.daffertshofer@vu.nl

Perception of speech and gestures engage common brain areas. Neural regions involved in speech perception overlap with those involved in speech production in an articulator-specific manner. Yet, it is unclear whether motor cortex also has a role in processing communicative actions like gesture and sign language. We asked whether the mere observation of hand gestures, paired and not paired with words, may result in changes in the excitability of the hand and tongue areas of motor cortex. Using single-pulse transcranial magnetic stimulation (TMS), we measured the motor excitability in tongue and hand areas of left primary motor cortex, while participants viewed video sequences of bimanual hand movements associated or not-associated with nouns. We found higher motor excitability in the tongue area during the presentation of meaningful gestures (noun-associated) as opposed to meaningless ones, while the excitability of hand motor area was not differentially affected by gesture observation. Our results let us argue that the observation of gestures associated with a word results in activation of articulatory motor network accompanying speech production.

Keywords: transcranial magnetic stimulation, tongue motor excitability, speech perception, gesture perception, sign language

INTRODUCTION

The processes underlying sign and spoken language perception are known to involve overlapping neural populations. Apparently linguistic information conveyed through gestures and sounds is processed in similar ways (Damasio et al., 1986; Hickok et al., 1996; Neville, 1998; MacSweeney et al., 2002; Newman et al., 2002; Xu et al., 2009; Straube et al., 2012, 2013). This overlap let McNeill (1996) speculate about a unified communication system.

It has been proposed that the evolutionary transition from gesticulation to speech has been mediated by the mirror neuron system, which is believed to underlie the understanding of others' actions and intentions (Rizzolatti and Arbib, 1998). Interestingly, mirror neurons have first been discovered in monkey area F5 that is considered homolog to human area 44 (Broca's area), which hosts speech production (Rizzolatti et al., 1996, 2002; Kohler et al., 2002). Nonetheless, no proper evidence supporting the evolution of language from gesture has emerged. According to this idea, vocal communication has become more and more autonomous at the expense of gestures that gradually lost their importance. In the *motor theory of speech perception*, Liberman and colleagues have already proposed the motor system to be involved in sensory perception (Liberman et al., 1967; Liberman and Mattingly, 1989). Hence, listeners may perceive spoken language by generating forward models in the motor system by activating articulatory phonetic gestures used to produce acoustic speech signals.

Imaging and transcranial magnetic stimulation (TMS) experiments revealed that speech perception triggers activity in brain areas that are involved in speech production in a somatotopic manner (Fadiga et al., 2002; Watkins et al., 2003; Pulvermüller et al., 2006; D'Ausilio et al., 2014; Möttönen et al., 2014). Repetitive TMS over the left premotor or primary motor cortex causes the capacity of phonetic discrimination to be significantly reduced (Meister et al., 2007; Möttönen and Watkins, 2009; Sato et al., 2009; Möttönen et al., 2014), indicating a causal relationship between the motor system and speech perception. Neural controllers of the articulator's movement seemingly contribute to both production and perception of speech. Nevertheless it has been argued that the activation of motor areas during listening to speech is neither essential in speech perception nor does it reflect phonetic processing of the speech signal as suggested in motor theory of speech perception (Scott et al., 2009). Evidence from functional lesion studies also supports the idea that involvement of motor areas during speech production does not necessarily contribute to speech perception (for review see, e.g., Hickok and Poeppel, 2000). Scott et al. (2009) argued that several different linguistic functions could be served by motor cortex during speech perception, including a specific role in sensorimotor processing in conversation. But is motor cortex activated differently during the observation of communicative actions such as gesture and sign language?

Recently, Möttönen et al. (2010) reported that motor evoked potentials (MEPs) elicited by stimulating the hand representation

in the primary motor cortex (M1) did not differ when participants observed signs with known vs. signs with unknown meanings. If M1 hand area seems insensitive to the distinction between action associated and not associated with words, then other regions in M1 like the tongue or lip areas might be better candidates for this (cf. Fadiga et al., 2002; Watkins et al., 2003; Roy et al., 2008; Sato et al., 2010; D'Ausilio et al., 2014). It remains unclear, however, whether the motor representations of tongue and lips are capable of distinguishing between those actions that symbolically represents words (e.g., an object or a state) and those that do not.

In this study we investigated whether observation of newly learned hand gestures paired and not paired with words may result in changes in the excitability of the hand and tongue areas of motor cortex. We studied MEPs recorded from tongue and hand muscles in a group of healthy Italian participants who had been taught some signs in American Sign Language (ASL). Participants were asked to observe signs associated and not associated with words, i.e., trained and untrained signs. We first trained participants to learn the associated words for several signs (through visual presentation of signs with the associated words as subtitle), while the associated words for the other half of the signs were not taught (the signs presented without subtitles). To ensure that all participants learned the associated words, they underwent a testing session during which participants were observing the video of all the signs but this time without subtitles. They were asked to choose corresponding words for the observed signs among four possible alternatives displayed on the screen. Finally, participants underwent a TMS session, during which we measured the motor excitability in tongue and hand areas of left primary motor cortex while participants were observing the stimuli. We expected the observation of hand gestures alone would lead to similar excitability of hand motor representation, regardless of whether they represent a word or not. We also expected that only the observation of hand gestures associated with words would modulate the excitability of the tongue motor representation.

MATERIALS AND METHODS

EXPERIMENTAL PROCEDURE

The experiment was designed as a 2×2 repeated measurement with two sign types (i.e., meaningful and meaningless indicating hand movements associated and not-associated with words, respectively) and two muscles (tongue and hand). During the experiment TMS-induced MEPs were recorded from tongue and hand muscles while participants observed video sequences of hand movements associated or not-associated with nouns. For each experimental condition 18 MEPs were recorded. The experiment was divided into three sessions: training, test, and TMS.

PARTICIPANTS

Ten non-signer adult, native Italian speakers (5 females; 23.5 ± 2.6) participated in the study. All were right-handed and had normal or corrected-to-normal vision with no history of speaking or hearing disorders. None of the participants were experienced in ASL. The experimental protocol had been approved

by the members of the Ethics Committee of the Department of Neurological, Neuropsychological, Morphological and Movement Sciences of the University of Verona. All participants provided their informed consent prior to entering the study, which had been approved by the institutional review board.

STIMULI

Stimuli consisted of six short (duration 3 s) black and white videos depicting hands performing bimanual movements (the actor's hands and trunk was presented against a gray background). The hand movements were signs in ASL, which were not related in movement structure to any Italian symbolic gestures. Moreover, ASL and not the Italian one was chosen to rule out the contingency of participants' familiarity with the signs used (having seen the signs and learned their related meanings). All the signs were nouns or adverbs (Necklace, Night, Land, Collision, Below, Current) consisting of double consonants "rr" "ll" "tt" in their Italian translation (Collana, Notte, Terre, Collisione, Sotto, Corrente), which require strong tongue mobilization for proper pronunciation; see **Figure 1**. The signs were chosen to share the following features: (1) contraction and visibility of the right hand first dorsal interosseous (FDI) muscle in the videos; (2) having associated words that require strong tongue mobilization when pronounced. The FDI muscle was visible and contracted in all the videos to reduce the variability among stimuli because previous studies showed that action observation under different circumstances may lead to modulation of corticospinal motor excitability (for a review, see Rizzolatti and Craighero, 2004). Further, to reduce the variability amongst stimuli concerning the associated words, which would share an element like the visibility and contraction of FDI muscle, we used words with double consonants. These words require strong tongue mobilization when produced, which has already been shown to modulate tongue motor excitability when listening to (Fadiga et al., 2002).

TRAINING SESSION

Before training participants were informed that they were going to view six videos of various hand movements each repeated ten times, three of which had related-word in the form of a subtitle and three did not. We restricted the study to six stimuli to ensure that all participants could readily learn the three associated words. Participants were instructed to memorize the association gesture-word from each of the three videos with the subtitles. One group learned the associated gesture-words of (Collana, Notte, Terre) and the other one learned (Collisione, Sotto, Corrente). The training was set up as follows: at first a screen-centered fixation cross was displayed for 1000 ms; subsequently, video stimuli (three with subtitles and three without) were presented in random order for 3 s. Participants were asked to be silent during the entire experiment. To test whether the participants learned the meaning of the signs, they underwent a test session after training.

TEST SESSION

During the test session the videos of the training session were presented at random without subtitles, each repeated ten times. After



FIGURE 1 | Illustration of the digital video clips presented to the subjects. All video clips had duration of 3 s.

each stimulus presentation, participants were asked to choose the corresponding word for the observed signs among four possible alternatives displayed on the screen until participants choose the correct answer by clicking the right mouse button with the index finger of the left hand. The four possible choices were the three learned words plus a question mark indicating “I do not know the answer”. The displays were centered in the four quadrants of the screen. For every answer participants received feedback of correctness (knowledge of results). The feedback for the correct answer was displayed in white on a black background in the center of screen, and the feedback for the incorrect answer was displayed in red on a white background. Stimulus order and target position on the screen were randomized. All the participants accomplished the test session successfully (100% of correct responses) without any errors rendering ongoing ASL learning unlikely.

TMS/EMG SESSION

Procedure

The experiment was designed using the E-Prime 2 (Psychology Software Tools, Inc, USA) software running on a PC computer with a Windows XP operating system to control the stimulus presentation, randomization of trials and to trigger the TMS and EMG recordings. Transcranial magnetic stimulation induced EMG activity was collected from all participants.

During the experiment, the subjects were comfortably seated on an armchair in a dimly-lit room at a distance of 80 cm from a computer screen (Asus, 17”, 60 Hz refresh rate). Each trial started with a fixation cue (the “+” symbol), presented for 1000 ms immediately followed by the stimulus that lasted for 3 s. The left M1 was stimulated via a single-pulse TMS delivered through a figure-of-eight coil at 120% of the individual resting motor threshold (over both tongue and hand motor areas). The TMS pulses were generated randomly within the last 2 s

of stimulus presentation (from the beginning of the second to the end of the third second), when in the observed actions the FDI muscle was contracting and the meaning had already been conveyed. This was done to ensure that FDI muscle was clearly observable when the TMS pulses are delivered, and to give the participants more time to recognize the associated word. After each stimulus presentation, participants were asked to choose the corresponding word for the observed signs in a same way they did during the testing session. After pressing the space button to continue, the next stimulus was delivered with an inter stimulus interval of 8 s. Every TMS session took about 15 min and consisted of 36 trials (18 per each condition). The two sessions (tongue and hand stimulation) were carried out on the same day and their order was counter-balanced across participants.

Data acquisition

Focal TMS was applied with a 70-mm figure-of-eight coil that was powered by a STM9000 Magnetic Stimulator (ATES Medical Device, IT) producing a maximum output of 2T at the coil surface. Before each session, the coil was moved over the scalp in order to determine the optimal site from which maximal amplitude MEPs were elicited in the tongue and hand muscles separately. The coil was held tangentially to the scalp with the handle pointing 45° away from the nasion–inion line in a posterolateral direction (Mills et al., 1992) to find the FDI representation area. Following the same procedure pursued in a previous work of our group (Vicario et al., 2014), the tongue area was stimulated with the coil handle oriented at 90° directed straight posteriorly.

The resting motor threshold of the muscles was determined according to standard methods as the minimal intensity capable of evoking MEPs in 5 out of 10 trials of the relaxed muscles with amplitude of at least 50 μ V (Rossini et al., 1994). Bipolar EMG

from the tongue muscles were acquired using a pair of Ag-AgCl surface electrodes (\varnothing 1 cm). The electrodes were pasted on plastic buttons and fixed on a spring of iron zinc. Before recording, electrodes were immersed in a disinfectant solution (Amuchina, sodium hypochlorite 1.1 grams per 100 ml of purified water) for 5 min and rinsed in drinking water. Participants were asked to introduce their tongue within these two electrodes, adjust the spring so that it was perfectly fitting with the tongue, and remain as relaxed as possible for the full duration of the experiment. The ground electrode was placed on the forehead of the participant. In separate sessions, EMG activity was recorded from the FDI muscle of the right hand by placing surface electrodes over the muscle belly (active electrode) and over the tendon of the muscle (reference electrode). The ground electrode was placed over the dorsal part of the elbow. The activity of muscles were registered in separate blocks and counterbalanced across participants. Electromyography signals were band-pass filtered online (20–3000 Hz), amplified (Digitimer, Hertfordshire, England) and sampled at a rate of 5 kHz (CED Micro 1401, Cambridge Electronic Design, Cambridge, England). Motor evoked potentials' peak-to-peak amplitude (in millivolts) were calculated off-line using Spike 2 (version 6, Cambridge Electronic Design) and stored on a computer. We determined muscle pre-activation through visual inspection and excluded contaminated trials from the analysis (6.3% of trials, See **Figure 2** for MEP examples).

STATISTICS AND RESULTS

Motor evoked potentials' amplitude values were normalized (*z*-scored) for every subject and muscle. A two-way repeated measures ANOVA was performed with two sign types (meaningful and meaningless) and two muscles (tongue and hand). *Post hoc* comparisons were performed by means of *t*-tests applying a Bonferroni correction for multiple comparisons when required. A partial-eta-squared statistic served as effect size estimate. The interaction between the sign types and muscles was significant; ($F_{(1,9)} = 7.875$, $p = 0.021$, $\eta^2 = 0.46$). Tongue cortical excitability was enhanced during the presentation of meaningful (trained) as compared to meaningless (untrained) signs ($p = 0.02$). That is, the presentation of word-associated gestures yielded an increase in tongue MEPs compared to the observation of signs that were not associated with words. The hand MEP *z*-scores did not reveal significant differences between the two types of signs (and, therefore, the mean *z*-scores were close to zero) ($p > 0.05$)¹. Further *post hoc* analysis (Bonferroni test) indicated that observation of word-associated signs elicited significantly larger MEP ampli-

tudes, relative to meaningless signs on the tongue compared to the FDI muscle ($p = 0.025$). By contrast, meaningless signs were accompanied by relative decrement of MEP amplitudes in *z*-scores on the tongue as compared to the FDI muscle ($p = 0.019$); see **Figure 3**. Moreover, the raw MEP amplitudes recorded from the hand for each individual participant were greater than those recorded from the tongue muscle (cf. **Figure 2**). Because MEPs amplitude values were normalized using *z*-scores for each muscle, differences between MEP amplitudes of the two muscles (as shown in the **Figure 3**) are not necessarily indicative of differences in the magnitude of excitability.

DISCUSSION

To the best of our knowledge we have provided the first experimental evidence for the modulation of excitability in the tongue area of M1 cortex as a function of observation of word-associated movements. We found the highest cortical excitability in the tongue area during the presentation of word-associated gestures compared with gestures not associated with any words (meaningless). On the contrary, the hand motor area presented the same level of excitability for both type of gestures. Our results are in line with the TMS study by Möttönen et al. (2010) showing that MEPs elicited by the stimulation of the hand representation in the left M1 did not significantly differ when participants observed signs with known vs. signs with unknown meanings. To unravel motor cortex modulation during gesture observation, they recorded TMS-induced MEPs from hand muscles of participants during sign language observation. They also compared the MEPs obtained before and after individuals learned the meanings of the signs presented and found that the excitability of left and right hand representation in M1 was equally lateralized before participants knew that the presented hand movements were signs. By contrast, after learning both known and unknown signs, the motor cortical excitability significantly increased only on the left M1 side, supporting the left hemispheric dominance for language processing (Knecht, 2000). Moreover, it has been suggested that brief inactivation of Broca's area by use of repetitive TMS affects verbal responses to gesture observation, suggesting the involvement of Broca's area in the instantaneous control of gestures and word pronunciation (Gentilucci et al., 2006). In addition, the very recent study by Vicario et al. (2013) showed that M1 might be indirectly involved in the mapping process of newly acquired, action-related, categorical associations.

The current work complements these findings and underscores the contribution of tongue but not hand motor area in the processing of communicative hand actions associated with the words. Several TMS studies have demonstrated modulation in the excitability of tongue motor area during speech perception (Fadiga et al., 2002; Watkins et al., 2003; Roy et al., 2008; Sato et al., 2010; D'Ausilio et al., 2014). It has been thoroughly argued that motor activation during speech perception emerges as a result of different task demands or experimental conditions rather than being an essential activity underlying speech perception (for review see, e.g., Lotto et al., 2009; Scott et al., 2009). Moreover, whether articulatory commands activated automatically and involuntarily during speech perception, is still

¹To test whether this non-significant result was due to a lack of statistical power, we conducted an analysis using G* Power (Faul et al., 2009) with *t*-test as family test and "Means: difference between two dependent means (matched pairs)" as a statistical test and "A priori: compute required sample size—given α , power, and effect size" as analysis type. The input parameters and their values were set as follows: Tail(s) = Two, Effect size $d_z = 0.3234$ (derived from differences in the mean (0.024) and the standard deviation (0.0742)), α err prob = 0.05 and Power ($1 - \beta$ err prob) = 0.8. This analysis indicated that the sample size ought to be increased to $N = 77$ for the hand muscle to reach statistical significance at a 0.05 level. It is hence unlikely that the non-significant results found for hand muscle could be attributed to a limited sample size.

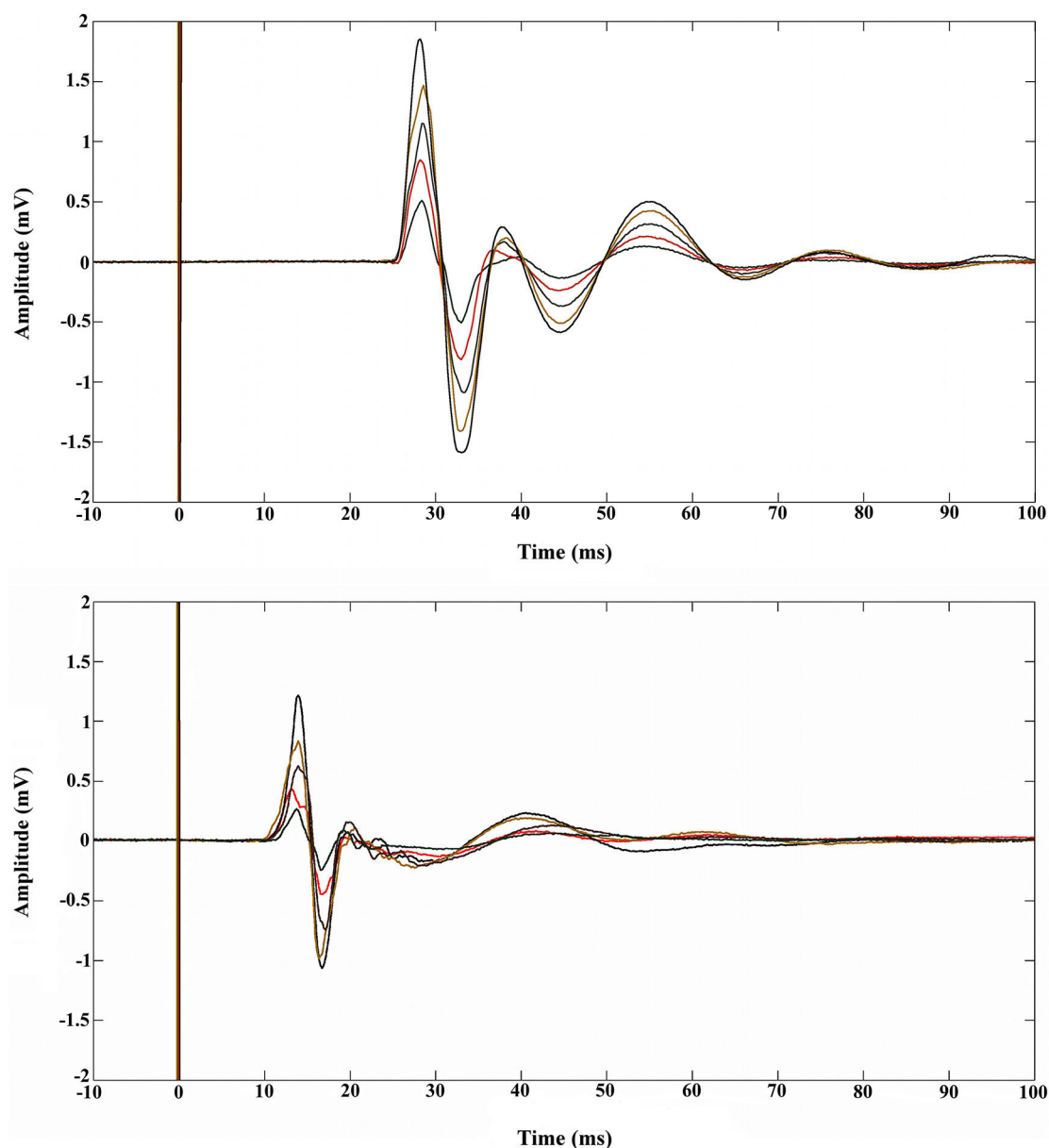
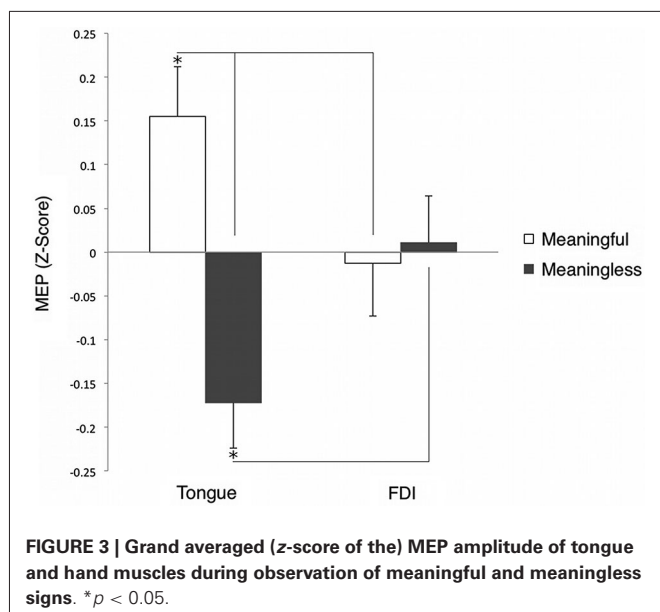


FIGURE 2 | Examples of five MEPs recorded during rest in the FDI muscle (upper panel) and in the tongue muscle (lower panel). The vertical lines at 0 ms indicate the moment at which the single pulse TMS was elicited.

a matter of debate (McGettigan et al., 2010). Here we have shown that it is not the hand but the tongue motor area that is specifically involved during the observation of gestures associated with a word, although individuals were not required to pronounce that word. Note that an additional control condition such as videos showing objects or symbols or fractals associated to specific words would have enabled us to determine whether the excitability in the M1 tongue area was a function of sign language observation or due to the effects of covert speech associated to the observed video. This should be addressed in future work.

Previous TMS studies have reported facilitation of the corticospinal tract excitability during the mere observation of another person's actions (for a review see Fadiga et al., 2005). The mirror system is active under various circumstances. For instance, somatotopic activation is present in the motor cortex when individuals observed and imagined actions (for a review, see Rizzolatti et al., 1996; Fadiga et al., 1999; Rizzolatti and Craighero, 2004). Even more critical is the mirror system involvement when the actions are not directly visible to the observer but implicitly presented (Bonaiuto et al., 2007). Building on to a vast amount of literature, one may speculate that humans have internal representations



of the movements either observed or imagined and that these internal representations resemble very closely the action when it is actually performed. In the present study we aimed for determining whether and how observation of hand gestures linked and not linked to specific words involves an internal motor simulation. We showed that while the observation of hand movements required similar internal motor simulations within the hand area of M1, regardless of whether they are associated with a word or not, only the observation of hand movements associated with words activated the tongue area of M1, indicating an extra level of coding. We here suggest that disentangling word-associated gestures, i.e., meaningful signs vs. meaningless signs, leads to internal simulations within the language motor regions (i.e., tongue).

The so-called gestural-origins theory of speech ascribes a precise role in language evolution to gestures (Corballis, 2003; see also Vicario, 2013 for a recent discussion). It has been suggested that spoken language evolves from an ancient communication system using arm gestures. Gestures of the mouth might have been added to the manual system to form a combined manuofacial gestural system (Corballis, 2003; Gentilucci and Corballis, 2006). Our results may suggest that the perception of sign language might require similar neural activity in speech motor centers as speech perception does. In this sense, our findings contribute to the view that the perception of speech and gesture share common neural substrates. Recent neuroimaging studies have revealed that semantic processing of speech and gestures engages common brain network with a specific involvement of left motor cortex (Xu et al., 2009; Straube et al., 2012, 2013). This is particularly interesting because it implies that motor cortex may be activated in response to language information independently of the communication modality.

Viewing sign language by deaf signers activates the classical language areas (left frontal and temporal areas) similar to the pattern of activity present when hearing participants listen to spoken words (Neville, 1998; MacSweeney et al., 2002; Newman

et al., 2002). Damage to the left hemisphere often produces sign language aphasia just like aphasia in spoken language, suggesting the left cerebral hemisphere dominance for both signed and spoken languages (Damasio et al., 1986; Hickok et al., 1996). Taken together we conclude that the involvement of the tongue region of the primary motor cortex is not merely limited to the perception and production of speech but might rather play a general role in encoding linguistic (maybe related to phonological retrieval) information even during perception of actions paired with words.

ACKNOWLEDGMENTS

We thank the Netherlands Organization for Scientific Research for financial support (NWO grant #400-08-127).

REFERENCES

- Bonaiuto, J., Rosta, E., and Arbib, M. (2007). Extending the mirror neuron system model, I. Audible actions and invisible grasps. *Biol. Cybern.* 96, 9–38. doi: 10.1007/s00422-006-0110-8
- Corballis, M. C. (2003). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Damasio, A., Bellugi, U., Damasio, H., Poizner, H., and Van Gilder, J. (1986). Sign language aphasia during left-hemisphere Amytal injection. *Nature* 322, 363–365. doi: 10.1038/322363a0
- D'Ausilio, A., Maffiongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., et al. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130418. doi: 10.1098/rstb.2013.0418
- Fadiga, L., Buccino, G., Craighero, L., Fogassi, L., Gallese, V., and Pavesi, G. (1999). Corticospinal excitability is specifically modulated by motor imagery: a magnetic stimulation study. *Neuropsychologia* 37, 147–158. doi: 10.1016/s0028-3932(98)00089-x
- Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402. doi: 10.1046/j.0953-816x.2001.01874.x
- Fadiga, L., Craighero, L., and Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Curr. Opin. Neurobiol.* 15, 213–218. doi: 10.1016/j.conb.2005.03.013
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/brm.41.4.1149
- Gentilucci, M., Bernardis, P., Crisi, G., and Dalla Volta, R. (2006). Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *J. Cogn. Neurosci.* 18, 1059–1074. doi: 10.1162/jocn.2006.18.7.1059
- Gentilucci, M., and Corballis, M. C. (2006). From manual gesture to speech: a gradual transition. *Neurosci. Biobehav. Rev.* 30, 949–960. doi: 10.1016/j.neubiorev.2006.02.004
- Hickok, G., Bellugi, U., and Klima, E. S. (1996). The neurobiology of sign language and its implications for the neural basis of language. *Nature* 381, 699–702. doi: 10.1038/381699a0
- Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138. doi: 10.1016/s1364-6613(00)01463-7
- Knecht, S. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain* 123, 2512–2518. doi: 10.1093/brain/123.12.2512
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848. doi: 10.1126/science.1070311
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279
- Lieberman, A. M., and Mattingly, I. G. (1989). A specialization for speech perception. *Science* 243, 489–494. doi: 10.1126/science.2740894

- Lotto, A. J., Hickok, G. S., and Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends Cogn. Sci.* 13, 110–114. doi: 10.1016/j.tics.2008.11.008
- MacSweeney, M., Woll, B., Campbell, R., Calvert, G. A., McGuire, P. K., David, A. S., et al. (2002). Neural correlates of British sign language comprehension: spatial processing demands of topographic language. *J. Cogn. Neurosci.* 14, 1064–1075. doi: 10.1162/089892902320474517
- McGettigan, C., Agnew, Z. K., and Scott, S. K. (2010). Are articulatory commands automatically and involuntarily activated during speech perception? *Proc. Natl. Acad. Sci. U S A* 107:E42. doi: 10.1073/pnas.1000186107
- McNeill, D. (1996). *Hand Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696. doi: 10.1016/j.cub.2007.08.064
- Mills, K. R., Boniface, S. J., and Schubert, M. (1992). Magnetic brain stimulation with a double coil: the importance of coil orientation. *Electroencephalogr. Clin. Neurophysiol.* 85, 17–21. doi: 10.1016/0168-5597(92)90096-t
- Möttönen, R., Farmer, H., and Watkins, K. E. (2010). Lateralization of motor excitability during observation of bimanual signs. *Neuropsychologia* 48, 3173–3177. doi: 10.1016/j.neuropsychologia.2010.06.033
- Möttönen, R., Rogers, J., and Watkins, K. E. (2014). Stimulating the lip motor cortex with transcranial magnetic stimulation. *J. Vis. Exp.* doi: 10.3791/51665
- Möttönen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825. doi: 10.1523/jneurosci.6018-08.2009
- Neville, H. J. (1998). Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proc. Natl. Acad. Sci. U S A* 95, 922–929. doi: 10.1073/pnas.95.3.922
- Newman, A. J., Bavelier, D., Corina, D., Jezard, P., and Neville, H. J. (2002). A critical period for right hemisphere recruitment in American Sign Language processing. *Nat. Neurosci.* 5, 76–80. doi: 10.1038/nn775
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U S A* 103, 7865–7870. doi: 10.1073/pnas.0509989103
- Rizzolatti, G., and Arbib, M. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194. doi: 10.1016/s0166-2236(98)01260-0
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* 3, 131–141. doi: 10.1016/0926-6410(95)00038-0
- Rizzolatti, G., Fogassi, L., and Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Curr. Opin. Neurobiol.* 12, 149–154. doi: 10.1016/s0959-4388(02)00308-2
- Rossini, P. M., Barker, A. T., Berardelli, A., Caramia, M. D., Caruso, G., Cracco, R. Q., et al. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalogr. Clin. Neurophysiol.* 91, 79–92. doi: 10.1016/0013-4694(94)90029-9
- Roy, A. C., Craighero, L., Fabbri-Destro, M., and Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *J. Physiol. Paris* 102, 101–105. doi: 10.1016/j.jphysparis.2008.03.006
- Sato, M., Buccino, G., Gentilucci, M., and Cattaneo, L. (2010). On the tip of the tongue: modulation of the primary motor cortex during audiovisual speech perception. *Speech Commun.* 52, 533–541. doi: 10.1016/j.specom.2009.12.004
- Sato, M., Tremblay, P., and Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang.* 111, 1–7. doi: 10.1016/j.bandl.2009.03.002
- Scott, S. K., McGettigan, C., and Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* 10, 295–302. doi: 10.1038/nrn2603
- Straube, B., Green, A., Weis, S., and Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: an fMRI study. *PLoS One* 7:e51207. doi: 10.1371/journal.pone.0051207
- Straube, B., He, Y., Steines, M., Gebhardt, H., Kircher, T., Sammer, G., et al. (2013). Supramodal neural processing of abstract information conveyed by speech and gesture. *Front. Behav. Neurosci.* 7:120. doi: 10.3389/fnbeh.2013.00120
- Vicario, C. M. (2013). FOXP2 gene and language development: the molecular substrate of the gestural-origin theory of speech? *Front. Behav. Neurosci.* 7:99. doi: 10.3389/fnbeh.2013.00099
- Vicario, C. M., Candidi, M., and Aglioti, S. M. (2013). Cortico-spinal embodiment of newly acquired, action-related semantic associations. *Brain Stimul.* 6, 952–958. doi: 10.1016/j.brs.2013.05.010
- Vicario, C. M., Komeilipoor, N., Cesari, P., Rafal, R. D., and Nitsche, M. A. (2014). Enhanced corticobulbar excitability in chronic smokers during visual exposure to cigarette smoking cues. *J. Psychiatry Neurosci.* 39, 232–238. doi: 10.1503/jpn.130086
- Watkins, K. E., Strafella, A. P., and Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994. doi: 10.1016/s0028-3932(02)00316-0
- Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., and Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proc. Natl. Acad. Sci. U S A* 106, 20664–20669. doi: 10.1073/pnas.0909197106

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 June 2014; accepted: 10 September 2014; published online: 30 September 2014.

Citation: Komeilipoor N, Vicario CM, Daffertshofer A and Cesari P (2014) Talking hands: tongue motor excitability during observation of hand gestures associated with words. *Front. Hum. Neurosci.* 8:767. doi: 10.3389/fnhum.2014.00767

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Komeilipoor, Vicario, Daffertshofer and Cesari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The neural processing of foreign-accented speech and its relationship to listener bias

Han-Gyol Yi¹, Rajka Smiljanic² and Bharath Chandrasekaran^{1,3*}

¹ SoundBrain Lab, Department of Communication Sciences and Disorders, Moody College of Communication, The University of Texas at Austin, Austin, TX, USA

² UT Sound Lab, Department of Linguistics, College of Liberal Arts, The University of Texas at Austin, Austin, TX, USA

³ Institute for Neuroscience, The University of Texas at Austin, Austin, TX, USA

Edited by:

Carolyn McGettigan, Royal Holloway
University of London, UK

Reviewed by:

Patti Adank, University College
London, UK

Daniel A. Abrams, Stanford
University, USA

*Correspondence:

Bharath Chandrasekaran,
SoundBrain Lab, Department of
Communication Sciences and
Disorders, Moody College of
Communication, The University of
Texas at Austin, 1 University Station,
C7000, 2504A Whitis Ave. (A1100),
Austin, TX 78712, USA
e-mail: bchandra@utexas.edu

Foreign-accented speech often presents a challenging listening condition. In addition to deviations from the target speech norms related to the inexperience of the nonnative speaker, listener characteristics may play a role in determining intelligibility levels. We have previously shown that an implicit visual bias for associating East Asian faces and foreignness predicts the listeners' perceptual ability to process Korean-accented English audiovisual speech (Yi et al., 2013). Here, we examine the neural mechanism underlying the influence of listener bias to foreign faces on speech perception. In a functional magnetic resonance imaging (fMRI) study, native English speakers listened to native- and Korean-accented English sentences, with or without faces. The participants' Asian-foreign association was measured using an implicit association test (IAT), conducted outside the scanner. We found that foreign-accented speech evoked greater activity in the bilateral primary auditory cortices and the inferior frontal gyri, potentially reflecting greater computational demand. Higher IAT scores, indicating greater bias, were associated with increased BOLD response to foreign-accented speech with faces in the primary auditory cortex, the early node for spectrotemporal analysis. We conclude the following: (1) foreign-accented speech perception places greater demand on the neural systems underlying speech perception; (2) face of the talker can exaggerate the perceived foreignness of foreign-accented speech; (3) implicit Asian-foreign association is associated with decreased neural efficiency in early spectrotemporal processing.

Keywords: foreign-accented speech, speech perception, fMRI, implicit association test, neural efficiency, primary auditory cortex, inferior frontal gyrus, inferior supramarginal gyrus

INTRODUCTION

Foreign-accented speech (FAS) can constitute an adverse listening condition (Mattys et al., 2012). Perception of FAS is often less accurate and more effortful compared to native-accented speech (NAS; Munro and Derwing, 1995b; Schmid and Yeni-Komshian, 1999; Van Wijngaarden, 2001). The reduced FAS intelligibility has been attributed to deviations from native speech in terms of segmental (Anderson-Hsieh et al., 1992; Van Wijngaarden, 2001) and suprasegmental (Anderson-Hsieh and Koehler, 1988; Anderson-Hsieh et al., 1992; Tajima et al., 1997; Munro and Derwing, 2001; Bradlow and Bent, 2008) cues. Nevertheless, listeners can adapt to FAS following exposure or training (Clarke and Garrett, 2004; Bradlow and Bent, 2008; Sidaras et al., 2009; Baese-Berk et al., 2013). Thus, listener's perception of FAS perception can improve over time (Bradlow and Pisoni, 1999; Bent and Bradlow, 2003). The neuroimaging literature on FAS perception is scant. However, perception of foreign phonemes has been shown to engage multiple neural regions. These include the superior temporal cortex, which matches the auditory input to the preexisting phonological representations ("signal-to-phonology mapping") in the articulatory network, encompassing the motor cortex, inferior frontal gyrus, and the insula (Golestani and

Zatorre, 2004; Wilson and Iacoboni, 2006; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). In particular, the inferior frontal gyrus exhibits phonetic category invariance, in which the response patterns differ according to between-category phonological variances but not to within-category acoustical variances (Myers et al., 2009; Rauschecker and Scott, 2009; Lee et al., 2012). Accordingly, processing of artificially distorted speech which, reduces speech intelligibility but does not necessarily introduce novel phonological representations, has been shown to involve additional recruitment of the superior temporal areas, the motor areas, and the insula, but not the inferior frontal gyrus (for review, see Adank, 2012). These findings lead to two predictions regarding neural activity during FAS processing. First, lack of adaptation to FAS would manifest in increased activity in the superior temporal auditory areas, due to the increased demand on auditory input processing. The primary auditory cortex is sensitive not only to rudimentary acoustic information such as frequency, intensity, and complexity of the auditory stimuli (Strainer et al., 1997), but also to the stochastic regularity in the input (Javit et al., 1994; Winkler et al., 2009) and attention (Jäncke et al., 1999; Fritz et al., 2003). The response patterns of the primary auditory cortex is modulated by task demands (attentional focus: Fritz

et al., 2003; target properties: Fritz et al., 2005), attention, training effects (frequency discrimination: Recanzone et al., 1993), and predictive regularity in the auditory input (Winkler et al., 2009). Furthermore, early acoustic signal processing time for speech stimuli has been shown to be reduced with accompanying visual information, indicating that the primary auditory cortex activity is modulated by crossmodal input (Van Wassenhove et al., 2005). Thus, the primary auditory cortex is attuned to analyzing the details of the incoming acoustic signals, but is also influenced by contextual information and modulated by experience. Second, difficulty in resolving phonological categories would manifest in increased activity in the articulatory network. In contrast to the early spectrotemporal analyses of speech, later stages of phonological processing are largely insensitive to within-category acoustic differences and exhibit enhanced sensitivity to across-category differences. Such phonological categorization is achieved via a complex network involving the inferior frontal cortex, insula, and the motor cortex (Myers et al., 2009; Lee et al., 2012; Chevillet et al., 2013).

Signal-to-phonology mapping, however, is not the only factor that modulates FAS perception. Listener beliefs regarding talker characteristics have been shown to modify the perceptual experience of speech (Campbell-Kibler, 2010; Drager, 2010). Specifically, different assumptions held about speaker properties by the individual listeners can potentially alter perception of the otherwise identical speech sounds. For instance, explicit talker labels (e.g., Canadian vs. Michigan) and indexical properties (e.g., gender, age, socioeconomic status) implied in visual representation of the talkers can change phonemic perception for otherwise identical speech sounds, even when the listeners are aware that this information is not accurate (Niedzielski, 1999; Strand, 1999; Hay et al., 2006a,b; Drager, 2011). The impact of perceived talker characteristics on FAS perception can be complex. Explicit labels have been linked to increased response times in lexical tasks for FAS, thus indicating increased processing load (Flocchia et al., 2009), while visual presentation of race-matched faces have been shown to increase intelligibility for Chinese-accented speech (McGowan, 2011). These findings suggest that listener variability in FAS intelligibility may be partly accounted for using measures of listeners' susceptibility to these indexical cues (Hay et al., 2006b). In social psychology, the implicit association test (IAT) has been used extensively to quantify the degree of implicit bias which may not be measured using explicit self-reported questionnaire entries (Greenwald et al., 1998, 2009; McConnell and Leibold, 2001; Bertrand et al., 2005; Devos and Banaji, 2005; Kinoshita and Peek-O'leary, 2005). During an IAT, the participants are instructed to make associations between two sets of stimuli (e.g., American vs. Foreign scenes; Caucasian vs. Asian faces). The response times between two conditions (e.g., Caucasian-American and Asian-Foreign vs. Caucasian-Foreign and Asian-American) are compared, and the magnitude of the difference between the mean RTs are considered to reflect the degree of implicit bias toward the corresponding association. In non-speech research domains, the IAT measures have been shown to be positively correlated with neural responses to dispreferred stimuli in various networks, including the amygdala, prefrontal cortex, thalamus, striatum, and the anterior cingulate cortex

(Richeson et al., 2003; Krendl et al., 2006; Luo et al., 2006; Suslow et al., 2010). A recent study has shown that native American English listeners with greater implicit bias toward making Asian-to-foreign and Caucasian-to-American associations experienced greater relative difficulties in transcribing English sentences in background noise, which were produced by native Korean speakers than that produced by native English speakers. This relationship between racial bias and FAS intelligibility was only observed when the auditory stimuli were paired with video recordings of the speakers producing the sentences (**Figure 1**; Yi et al., 2013). In spite of the novelty of the finding, did not reach a conclusive implication of the behavioral results, but rather cautiously suggesting that the listener bias likely led to altered incorporation of visual cues which are beneficial for enhancing speech intelligibility in adverse listening situations (Sumby and Pollack, 1954; Grant and Seitz, 2000). The precise neural mechanism underlying the relationship between listener bias and FAS perception remains unclear.

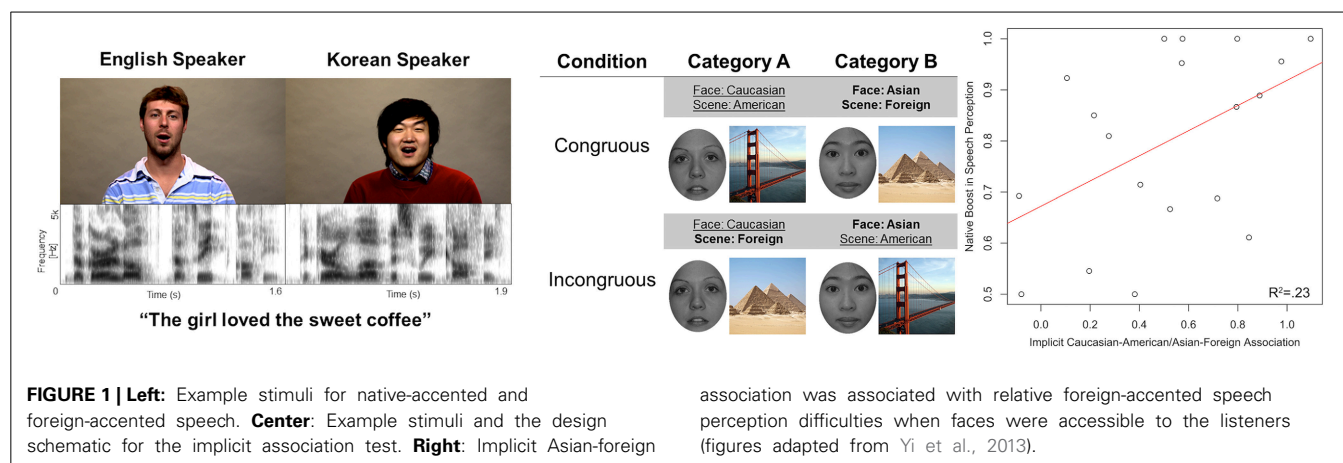
In this fMRI study, monolingual native English speakers ($N = 19$) were presented with English sentences produced by native English or native Korean speakers in an MR scanner. The sentences were presented either along with video recordings of the speakers producing the sentences ("audiovisual modality") or without ("audio-only modality"). A rapid event-related design was used to acquire functional images. This setup allowed us to independently estimate BOLD responses to stimulus presentation and motor response on a trial-by-trial basis. Outside the scanner, the participants performed an IAT which was designed to measure the extent of the association between Asian faces and foreignness. Whole brain analyses were conducted to test the prediction that FAS perception would involve increased activation in the superior temporal cortex and the articulatory-phonological network, consistent with previous research on foreign phonemes processing (Golestani and Zatorre, 2004; Wilson and Iacoboni, 2006), speech intelligibility processing (Adank, 2012), and categorical perception in the inferior frontal gyrus (Myers et al., 2009; Rauschecker and Scott, 2009; Lee et al., 2012). ROI analyses were conducted to test whether the degree of implicit association between Asian faces and foreignness would be associated with modifications in the signal-to-phonology mapping process. For this purpose, the ROI analysis was restricted to the primary auditory cortex, involved in auditory input processing, and the inferior frontal gyrus, involved in phonological processing. Previous neuroimaging studies utilizing IAT as a covariate have consistently shown positive correlation between the IAT scores and the neural response to the dispreferred stimuli, which has led us to hypothesize that higher IAT scores (stronger Asian-Foreign and Caucasian-American association) would be associated with greater BOLD response to FAS, especially in the audiovisual modality.

MATERIALS AND METHODS

MATERIALS

Participants

Nineteen young adults (age range: 18–35; 11 female) were recruited from the Austin community. All participants passed a hearing-screening exam (audiological thresholds <25 dB HL across octaves from 500 to 4000 Hz), had normal or corrected



association was associated with relative foreign-accented speech perception difficulties when faces were accessible to the listeners (figures adapted from Yi et al., 2013).

to normal vision, and self-reported to be right-handed. Potential participants were excluded if their responses to a standardized language history questionnaire revealed significant exposure to any language other than American English (LEAP-Q; Marian et al., 2007). Data from one participant (male) were excluded from all analysis due to detection of a structural anomaly. All recruitment and participation procedures were conducted in adherence to the University of Texas at Austin Institutional Review Board.

Audiovisual speech stimuli

Four native American English (2 female) and four native Korean speakers (2 female) produced 80 meaningful sentences with four keywords each (e.g., “the GIRL LOVED the SWEET COFFEE”; Calandruccio and Smiljanic, 2012). The speakers were between 18 and 35 years of age. The speakers were instructed to read text provided on the prompter as if they were talking to someone familiar with their voice and speech patterns. The NAS stimuli had been rated to be 96.2% native-like, and the FAS stimuli had been rated to be 20.7% native-like (converted from a 1-to-9 Likert scale; Yi et al., 2013). Twenty non-overlapping sentences from each speaker was selected, resulting in 80 sentence stimuli used in the experiment. The video track was recorded using a Sony PMW-EX3 studio camera, and the audio track was recorded with an Audio Technica AT835b shotgun microphone placed on a floor stand in front of the speaker. Camera output was processed through a Ross crosspoint video switcher and recorded on an AJA Pro video recorder. The recording session was conducted on a sound-attenuated sound stage at The University of Texas at Austin. The raw video stream was exported using the following specifications. Codec: DV Video (dvsd); resolution: 720 × 576; frame rate: 29.969730 (Figure 1). The raw audio stream was RMS amplitude normalized to 62 dB SNL and exported using the following specifications. Codec: PCM S16 LE (araw); mono; sample rate: 48 kHz; 16 bits per sample.

IAT

Ten young adult Asian (5 female) and 10 Caucasian (5 female) face images were used for Caucasian vs. Asian face categories (Minear and Park, 2004). All face images had been edited to

exclude hair, face contour, ear, and neck information, then rendered into grayscale with constant luminosity (Goh et al., 2010). Public domain images of 10 iconic American scenes (Grand Canyon, Statue of Liberty, Wrigley Field, Golden Gate Bridge, Pentagon, Liberty Bell, White House, Capitol, New York Central Park, Empire State Building) and 10 non-American foreign scenes (Eiffel Tower, Pyramids, Angkor Wat, London Bridge, Brandenburg Gate, Stonehenge, Great Wall of China, Leaning Tower of Pisa, Sydney Opera House, Taj Mahal) were obtained online and used for American vs. Foreign scene categories. No scene image contained face information. All images were cropped to a square proportion. The stimuli and the design used for the IAT were identical to those used in our previous study (Figure 1; Yi et al., 2013).

METHODS

Scan parameters

The participants were scanned via the Siemens Magnetom Skyra 3T MRI scanner at the Imaging Research Center of the University of Texas at Austin. High-resolution whole-brain T1-weighted anatomical images were obtained via MPRAGE sequence ($TR = 2.53$ s; $TE = 3.37$ ms; $FOV = 25$ cm; 256×256 matrix; 1×1 mm voxels; 176 axial slices; slice thickness = 1 mm; distance factor = 0%). T2*-weighted Whole-brain blood oxygen level dependent (BOLD) images were obtained using a gradient-echo multi-band EPI pulse sequence (flip angle = 60° ; $TR = 1.8$ s; 166 repetitions; $TE = 30$ ms; $FOV = 25$ cm; 128×128 matrix; 2×2 mm voxels; 36 axial slices; slice thickness = 2 mm; distance factor = 50%) using GRAPPA with an acceleration factor of 2. Three hundred and thirty-four time points were collected, resulting in the scanning duration of approximately 10 min. This was a part of a larger scanning protocol which lasted for approximately 1 h for each participant.

fMRI task

Participants were instructed to listen to the recorded sentences and rate the clarity of each one. After the presentation of each stimulus, a screen prompting the response was presented, upon which the participants rated the clarity of the stimulus by pressing one of the four buttons on the button boxes, ranging from

1 (not clear) to 4 (very clear). This was done to ensure that participants were attending to the presentation of the stimuli. The audio track for the sentences were presented auditorily via MR-compatible insert earphones (ER30; Etymotic Research), and the visual track was presented via projector visible by an in-scanner mirror. The stimuli were spoken by native English or native Korean speakers. There were two experimental conditions: an audio-only condition where only the acoustic signals were presented with a fixation cross being displayed, and an audio-visual condition where the video recordings of the talkers' faces producing the sentences were also presented. All sentences were presented only once in a single session without breaks. Therefore, the 80 sentences were subdivided into 20 sentences per each of the four conditions (native with visual cues; native without visual cues; nonnative with visual cues; nonnative without visual cues). We used a rapid event related design with jittered interstimulus intervals of 2–3 s. The order of the stimuli followed a pseudorandom sequence predetermined to avoid consecutive runs of stimuli of a given condition.

IAT

Following the fMRI acquisition session, IAT was conducted outside the scanner in a soundproof testing room. The IAT procedures were identical to those used in our previous study (Yi et al., 2013). For each trial, a face or scene stimulus was displayed on the screen. The face stimuli differed from the main task in the scanner in that they were still images unrelated to sentence production. In the congruous category condition, participants had to press a key on the keyboard when they saw a Caucasian face or an American scene, and a different key for an Asian face or a Foreign scene. In the incongruous category condition, participants had to press a key for a Caucasian face or a Foreign scene, and a different key for an Asian face or an American scene (Devos and Banaji, 2005). Participants were instructed to respond as quickly as possible without sacrificing accuracy. Each condition was presented twice with the key designations switched in a randomized order. These yielded four test blocks. Four practice blocks were included prior to the test blocks, in which only scenes or faces were presented. An incorrect response led to a corrective feedback of "Error!" (Greenwald et al., 2003).

ANALYSES

IAT

In a standalone analysis, a linear mixed effects analysis (Bates et al., 2012) was run with the response times in milliseconds as the dependent variable to directly quantify the delayed response times due to the incongruous association. The category condition (congruous vs. incongruous) and the neural index were entered as the fixed effects to measure the delay effect of face-scene pairings incongruent with the implicit association. By-subject random intercepts were included. The optimizer was set to BOBYQA (Powell, 2009). Individual IAT scores were calculated following the standard guidelines (Greenwald et al., 2003). Trials with response times longer than 10,000 ms or shorter than 400 ms were excluded. Response times for incorrect trials were replaced by the mean of the response times for correct trials within the same block, increased by 600 ms. The average response time discrepancies across the two pairs of congruous vs. incongruous

blocks were divided by the standard deviation of response times in the two blocks. These two discrepancy measures were averaged to yield in the final IAT score, which was used as a covariate in other analyses.

Clarity rating

Clarity ratings for all sentences from each participant were entered as the dependent variable, after being mean-centered to 0, in a linear mixed effects analysis. In order to counteract different clarity criteria across the participants, the model was corrected for by-participant random intercepts. In the first analysis, the fixed effects included the accent and modality of the stimuli, the individual IAT measures, and the ensuing interactions. The optimizer was set to BOBYQA (Powell, 2009).

fMRI preprocessing

fMRI data were analyzed using FMRIB's Software Library Version 5.0 (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012). BOLD images were motion corrected using MCFLIRT (Jenkinson et al., 2002). All images were brain-extracted using BET (Smith, 2002; Jenkinson et al., 2005). Registration to the high-resolution anatomical image ($df = 6$) and the MNI 152 template ($df = 12$; Grabner et al., 2006a) was conducted using FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002). Six separate block-wise first-level analysis were run within-subject. The following pre-statistics processing were applied; spatial smoothing using a Gaussian kernel (FWHM = 5 mm); grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0 s). Each event was modeled as an impulse convolved with a canonical double-gamma hemodynamic response function (phase = 0 s). Motion estimates were modeled as nuisance covariates. Temporal derivative of each event regressor, including the motion estimates, was added. Time-series statistical analysis was carried out using FILM with local autocorrelation correction (Smith et al., 2004). The event regressors consisted of stimulus, response screen, and clarity response. The stimulus regressors were subdivided into accent (native vs. foreign) and modality (audiovisual vs. audio-only) conditions. The missed trials were separately estimated as nuisance variables. Three sets of *t*-test contrast pairs were tested, which examined modality (audiovisual – audio-only; audio-only – audiovisual), accent (native-accented – foreign-accented; foreign-accented – native-accented), and the interaction effects (audiovisual native – audiovisual foreign – audio-only native + audio-only foreign; audiovisual foreign – audiovisual native – audio-only foreign + audio-only native).

Whole brain analysis

Group analysis was performed for each contrast using FLAME1 (Woolrich et al., 2009). To correct for multiple comparisons, post-statistical analysis was performed using randomize in FSL to run permutation tests ($n = 50,000$) for the GLM and yield in threshold-free cluster enhancement (TFCE) estimates of statistical significance. The corresponding family-wise error corrected *p*-values are presented in the results (Freedman and Lane, 1983; Kennedy, 1995; Bullmore et al., 1999; Anderson and Robinson,

Table 1 | Whole brain analysis results for the contrasts of interest.

Contrast	Regions	x (mm)	y (mm)	z (mm)	Voxels
MODALITY EFFECT					
(a) Audiovisual – audio-only	Bilateral occipital cortex; Bilateral fusiform gyri; L posterior superior temporal gyrus; Bilateral posterior middle temporal gyri	38	–46	–22	14108
	R thalamus	16	–32	2	208
	L thalamus	–26	–30	–4	188
	R amygdala	20	–4	–14	16
	R temporal pole	56	6	–22	13
(b) Audio-only – audiovisual	R superior parietal lobule;	34	–38	50	1864
	R somatosensory cortex;				
	R supramarginal gyrus				
	R superior frontal gyrus;	24	4	64	804
	R primary motor cortex				
	L superior parietal lobule;	–38	–36	42	330
	R somatosensory cortex				
	L middle frontal gyrus	–32	42	30	315
	L superior frontal gyrus	–26	4	56	177
	R middle frontal gyrus	30	36	22	82
ACCENT EFFECT					
(c) Native – foreign	R posterior middle temporal gyrus;	64	–46	26	379
	R posterior inferior temporal gyrus;				
	R angular gyrus; R supramarginal gyrus				
(d) Foreign – native	Paracingulate gyrus	4	24	34	904
	R motor cortex, R superior parietal lobule;	34	–52	62	868
	R somatosensory cortex				
	L insular cortex	–40	14	8	76
	R superior frontal gyrus	20	–2	60	68
	R insular cortex	32	20	–6	50
	L inferior frontal gyrus	–52	10	10	42
	L insular cortex	–26	24	0	31
	R insular cortex	42	14	8	13
Modality by accent interaction	n.s.				

Clusters are based on the $p < 0.025$ threshold as well as the size criterion of 10 voxels.

2001; Nichols and Holmes, 2002; Hayasaka and Nichols, 2003). The results are presented in the **Table 1**.

ROI analysis

The ROIs were anatomically defined as the left and right primary auditory cortices (combination of Te 1.0, 1.1, and 1.2; Morosan et al., 2001) and the left inferior frontal gyrus (Brodmann area 44; Amunts et al., 1999) using the Jülich histological atlas (threshold = 25%; Eickhoff et al., 2005, 2006, 2007). Percent changes in BOLD responses for the stimuli in four conditions (native-accented with faces; native-accented without faces; foreign-accented with faces; foreign-accented without faces) were calculated by first linearly registering the ROIs to the individual BOLD spaces using FLIRT with the appropriate transformation matrices generated from the first level analysis and nearest neighbor interpolation (Jenkinson and Smith, 2001; Jenkinson et al., 2002). Then, the parameter estimate images were masked for the transformed ROIs, multiplied by height of the double

gamma function for the stimulus length of 2 s (0.4075), converted into percent scale, divided by mean functional activation, and averaged within the ROI, using fslmaths (Mumford, 2007). The percent signal change was entered as the dependent variable in a linear mixed effects analysis. In the mixed effects analysis, the fixed effects included the accent (native vs. foreign), face (faces vs. no faces), individual IAT values and their interaction terms. The model was corrected for by-participant random intercepts (Bates et al., 2012). The optimizer was set to BOBYQA (Powell, 2009).

RESULTS

BEHAVIORAL RESULTS

Clarity ratings

The overall mean clarity rating was 2.94 ($SD = 1.09$). The mean clarity rating for the NAS was 3.28 ($SD = 1.10$) in the audio-only condition and 3.34 ($SD = 1.06$) in the audiovisual condition, while the rating for the FAS was 2.55 ($SD = 0.96$) in the audio-only condition and 2.57 ($SD = 0.96$) in the audiovisual

condition. In this analysis, the fixed effects of modality, accent, the IAT scores, and their interaction terms were included as fixed effects for the dependent variable of clarity ratings for each sentence, which was mean-centered to 0. The three-way interaction was not significant, which was excluded in the final model. The accent effect was significant, $b = -0.89$, $SE = 0.095$, $t = -9.36$, $p < 0.0001$, 95% CI $[-1.08, -0.70]$, indicating that FAS was perceived to be less clear than NAS. The accent by IAT interaction was significant, $b = 0.33$, $SE = 0.15$, $t = 2.13$, $p = 0.034$, 95% CI $[0.026, 0.63]$, indicating that higher IAT values were more associated with higher perceived clarity ratings for the FAS relative to NAS. The intercept was significant, $b = 0.97$, $SE = 0.39$, $t = 2.45$, $p = 0.024$, 95% CI $[0.15, 1.78]$. The modality effect was not significant, $b = -0.015$, $SE = 0.094$, $t = -0.16$, $p = 0.88$, 95% CI $[-0.20, 0.17]$, failing to provide evidence that perceived clarity was modified by the availability of visual cues. This is in contrast to the extensive previous literature that have indicated the intelligibility benefit from the audiovisual modality (Sumbly and Pollack, 1954; Macleod and Summerfield, 1987; Ross et al., 2007). We ascribe this null finding to the task properties which did not require the participants to actively decipher the sentences, but only to rate their clarity (Munro and Derwing, 1995a). The IAT effect was not significant, $b = -0.34$, $SE = 0.70$, $t = -0.49$, $p = 0.63$, 95% CI $[-1.79, 1.10]$. The modality by accent interaction was not significant, $b = -0.054$, $SE = 0.077$, $t = -0.71$, $p = 0.48$, 95% CI $[-0.20, 0.096]$. The modality by IAT interaction was not significant, $b = 0.15$, $SE = 0.15$, $t = 0.97$, $p = 0.33$, 95% CI $[-0.15, 0.45]$. These results altogether suggest that FAS is perceived to be less clear by the listeners. Participants with higher IAT scores, i.e., those who were more likely to implicitly associate East Asian faces with foreignness, have decreased tendency to perceive FAS to be unclear, compared to NAS.

IAT

The overall mean response time was 948 ms ($SD = 586$ ms). The mean RT was 824 ms ($SD = 408$ ms) in the congruous condition, and 1073 ms ($SD = 700$ ms) in the incongruous condition. One fixed effects term was included in the model: incongruity of the

stimuli pairing. The intercept was significant, $b = 823.83$, $SE = 50.92$, $t = 19.50$, $p < 0.0001$, 95% CI $[718.84, 928.81]$, showing approximately 820 ms baseline response time. The incongruity effect was significant, $b = 249.24$, $SE = 19.90$, $t = 12.52$, $p < 0.0001$, 95% CI $[210.22, 288.27]$, suggesting that incongruous stimuli pairing delayed each response by approximately 250 ms. The mean IAT score was calculated to be 0.51 ($SD = 0.25$), indicating a general trend of implicit bias toward making the Asian-Foreign association.

fMRI RESULTS

Audio-only vs. Audiovisual

BOLD signals were compared across the audiovisual and audio-only stimuli. The [audiovisual – audio-only] contrast revealed extensive activity in the occipital cortex, as the visual information in the faces required computations in the visual modality. Activity in the bilateral middle temporal gyri, left posterior superior temporal gyrus, and the right temporal pole was also observed, presumed to reflect integrative effort of the visual cues available in the facial stimuli (Sams et al., 1991; Möttönen et al., 2002; Pekkola et al., 2005). The [audio-only – audiovisual] contrast revealed activity in the bilateral superior and middle frontal gyri, right motor and somatosensory cortices, and the bilateral supramarginal gyri (Figure 2). The increased activation in the motor and somatosensory areas for audio-only speech than for audiovisual speech is in contrast to previous research that has shown the opposite pattern (Skipper et al., 2005). It is possible that the absence of visual cues induced more effortful processing in these areas. The activity in these regions is presumed to reflect the necessity of additional computation in the speech processing network.

Native- vs. Foreign-accented speech

BOLD signals were compared across the speaker accent. The [native – foreign] contrast revealed greater activity in the right angular gyrus, supramarginal gyrus, the posterior middle, and inferior temporal gyri. Supramarginal gyri have been suggested to be involved in making phonological decisions, which in the context of this study is presumed to reflect improved phonological

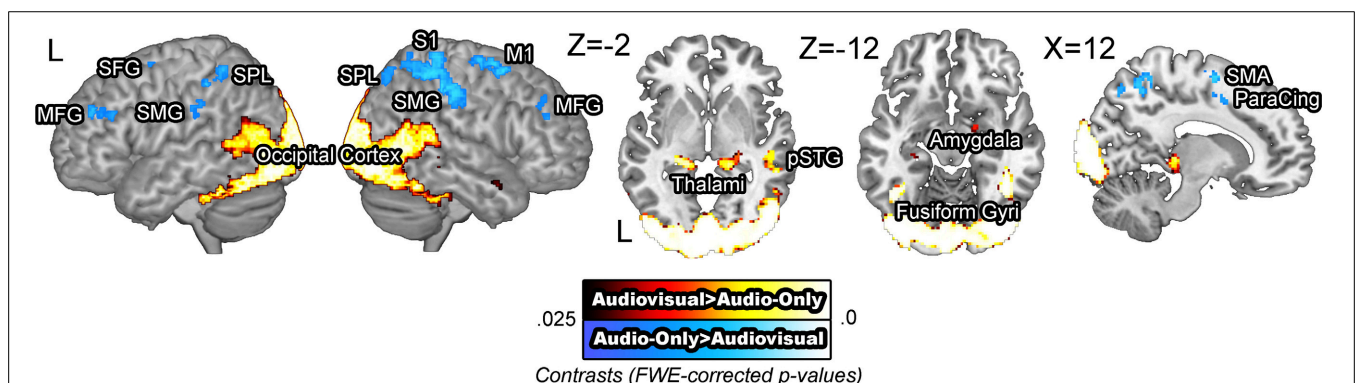
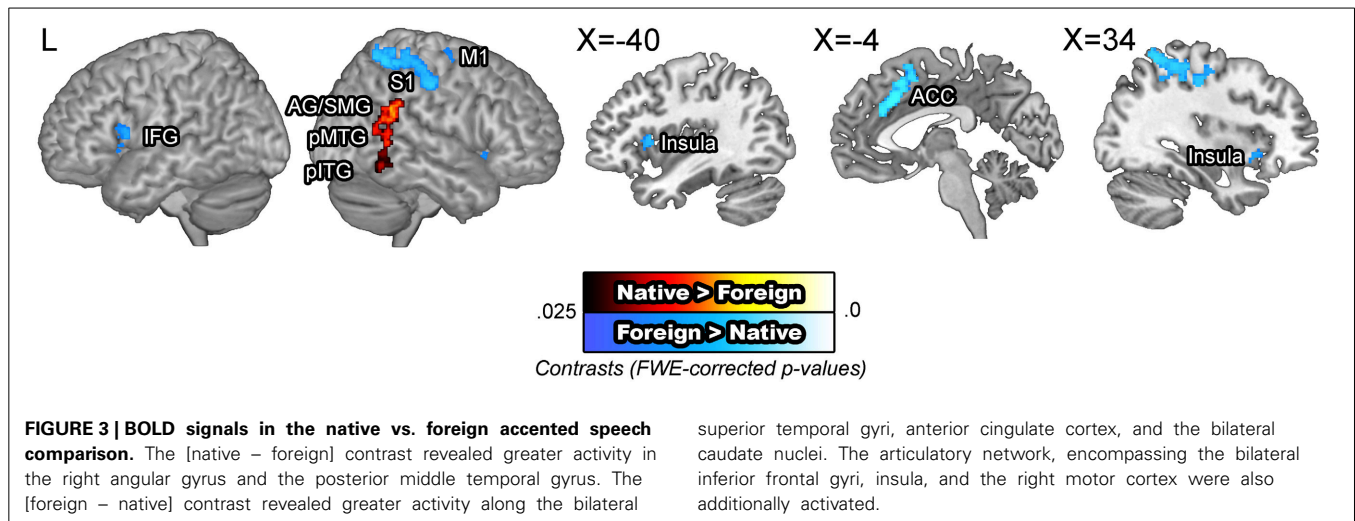


FIGURE 2 | BOLD signals in the audio-only vs. audiovisual comparison.

The [audiovisual – audio-only] contrast revealed extensive activity in the occipital cortex and the bilateral posterior middle temporal gyri. The

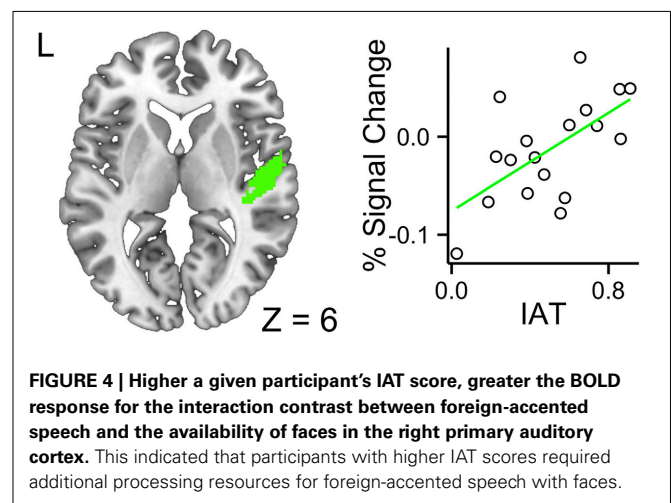
[audio-only – audiovisual] contrast revealed activity in the right middle frontal gyrus, right motor and somatosensory areas and the superior parietal lobule.



processing for NAS than for FAS (Hartwigsen et al., 2010). The [foreign – native] contrast revealed greater activity in the motor cortex, somatosensory cortex, inferior frontal gyrus, insula, and the anterior cingulate cortex. These areas have been previously indicated to be additionally recruited for perception of foreign phonemes (Golestani and Zatorre, 2004; Wilson and Iacoboni, 2006) or distorted speech (Adank, 2012). The omission of the superior temporal areas are significant, which run counter to our initial hypothesis regarding increased computational demand due to unfamiliar auditory input. A potential interpretation of this null finding is that the activity in the superior temporal cortex was more variable than that in the motor and frontal areas, an idea which was tested in the subsequent ROI analysis (Figure 3). The modality by accent interaction contrasts did not yield significant results.

ROI analyses

The ROI analyses were constrained to the left and right primary auditory cortices and the left inferior frontal gyrus. The fixed effects included accent (foreign- vs. native-accented speech), modality (audiovisual vs. audio-only), IAT scores, and their interaction terms. In the left primary auditory cortex, no three-way or two-way interactions were significant, leaving the model with only three main effects of accent, modality and IAT to be considered. The modality effect was significant, $b = -0.047$, $SE = 0.014$, $t = -3.34$, $p = 0.0016$, 95% CI $[-0.075, -0.020]$, suggesting that the audiovisual stimuli reduced computational demand in this region, relative to the audio-only stimuli. The accent effect was not significant, $b = 0.010$, $SE = 0.014$, $t = 0.71$, $p = 0.48$, 95% CI $[-0.018, 0.038]$. The IAT effect was not significant, $b = -0.27$, $SE = 0.20$, $t = -1.31$, $p = 0.21$, 95% CI $[-0.66, 0.13]$. The intercept was significant, $b = 0.45$, $SE = 0.11$, $t = 3.89$, $p = 0.0013$, 95% CI $[0.22, 0.67]$. In the right primary auditory cortex, the three-way interaction across modality, accent, and IAT was significant, $b = 0.25$, $SE = 0.11$, $t = 2.26$, $p = 0.028$, 95% CI $[0.042, 0.46]$, suggesting that higher IAT scores were associated with increased response to FAS with faces (Figure 4). The interaction between accent and modality was



significant, $b = -0.15$, $SE = 0.062$, $t = -2.43$, $p = 0.019$, 95% CI $[-0.27, -0.34]$, suggesting that the decreased neural efficiency due to FAS was ameliorated by the availability of faces. The accent effect was significant, $b = 0.090$, $SE = 0.044$, $t = 2.04$, $p = 0.047$, 95% CI $[0.0070, 0.17]$, suggesting that FAS increased the computational demand in this region. The intercept was significant, $b = 0.32$, $SE = 0.12$, $t = 2.57$, $p = 0.020$, 95% CI $[0.077, 0.55]$. The accent by IAT interaction was not significant, $b = -0.061$, $SE = 0.078$, $t = -0.78$, $p = 0.44$, 95% CI $[-0.21, 0.087]$. The IAT by modality interaction was not significant, $b = -0.13$, $SE = 0.078$, $t = -1.65$, $p = 0.11$, 95% CI $[-0.28, 0.018]$. The modality by accent interaction was not significant, $b = -0.15$, $SE = 0.062$, $t = -2.43$, $p = 0.019$, 95% CI $[-0.27, -0.034]$. The IAT main effect was not significant, $b = -0.23$, $SE = 0.22$, $t = -1.03$, $p = 0.32$, 95% CI $[-0.65, 0.20]$. The modality main effect was not significant, $b = 0.011$, $SE = 0.044$, $t = 0.25$, $p = 0.81$, 95% CI $[-0.072, 0.094]$. The intercept was significant, $b = 0.32$, $SE = 0.12$, $t = 2.57$, $p = 0.020$, 95% CI $[0.077, 0.55]$.

In the left inferior frontal gyrus, no three-way or two-way interactions were significant, leaving the model with only three main effects of accent, modality and IAT to be considered. The accent effect was significant, $b = 0.058$, $SE = 0.018$, $t = 3.15$, $p = 0.0028$, 95% CI [0.022, 0.094], suggesting that FAS increased computational demand in this region. The IAT effect was not significant, $b = -0.37$, $SE = 0.20$, $t = -1.83$, $p = 0.086$, 95% CI [-0.76, 0.024]. The modality effect was not significant, $b = -0.026$, $SE = 0.018$, $t = -1.41$, $p = 0.16$, 95% CI [-0.062, 0.010]. The intercept was not significant, $b = 0.12$, $SE = 0.11$, $t = 1.02$, $p = 0.32$, 95% CI [-0.11, 0.34]. In the right inferior frontal gyrus, no three-way or two-way interactions were significant, leaving the model with only three main effects of accent, modality, and IAT to be considered. The accent effect was not significant, $b = 0.0080$, $SE = 0.019$, $t = 0.43$, $p = 0.67$, 95% CI [-0.029, 0.045]. The IAT effect was not significant, $b = -0.36$, $SE = 0.19$, $t = -1.89$, $p = 0.077$, 95% CI [-0.74, 0.011]. The modality effect was not significant, $b = -0.0090$, $SE = 0.019$, $t = -0.48$, $p = 0.63$, 95% CI [-0.046, 0.028]. The intercept was not significant, $b = 0.12$, $SE = 0.11$, $t = 1.10$, $p = 0.29$, 95% CI [-0.093, 0.33].

DISCUSSION

Listening to FAS can be challenging compared to NAS. This difficulty can be partly attributed to a demanding process of mapping somewhat unreliable incoming signals to phonology. We hypothesized that FAS perception will require additional spectrotemporal analysis of the acoustic signal and place a greater demand on the phonological processing network. We therefore predicted increased functional activity in the superior temporal cortex and the inferior frontal gyrus, insula, and the motor cortex (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Adank and Devlin, 2010; Adank et al., 2012a,b). Furthermore, FAS perception is additionally modulated by listeners' underlying implicit bias (Greenwald et al., 1998, 2009; McConnell and Leibold, 2001; Bertrand et al., 2005; Devos and Banaji, 2005; Kinoshita and Peek-O'leary, 2005; Yi et al., 2013). Thus, we hypothesized that individual variability in implicit Asian-foreign association will be associated with functional activity during early spectrotemporal analysis in the primary auditory cortex or for later, more categorical processing in the inferior frontal gyrus (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009).

INCREASED COMPUTATIONAL DEMAND FOR FOREIGN-ACCENTED SPEECH PROCESSING

Relative to native speech, FAS was associated with increased BOLD response in the bilateral superior temporal cortices, potentially reflecting increased computational demand on these regions. The anterior and posterior portions of the superior temporal cortex have been associated with spectrotemporal analysis of the speech signal (Hickok and Poeppel, 2007), as well as with speech intelligibility processing (Scott et al., 2000; Narain et al., 2003; Okada et al., 2010; Abrams et al., 2013). While these previous studies have observed increased activation of the superior temporal cortex for intelligible speech compared to unintelligible acoustically complex stimuli, we found increased activation for the FAS stimuli than for the NAS stimuli, although FAS is

less intelligible than NAS (Yi et al., 2013). This contradiction is resolved by considering the nature of comparisons involved in previous neuroimaging studies examining mechanisms underlying speech intelligibility. Both native- and FAS used in the current study have semantic and syntactic content which are absent in non-speech stimuli used as control in the previous studies (e.g., spectrally-rotated speech), and both functions have been suggested to occur within the superior temporal cortex (Friederici et al., 2003). The superior temporal cortex is a large region with possibly multiple functional roles in processing information in the speech signal. A direct recording study has shown that speech sound categorization is represented in the posterior superior temporal cortex (Chang et al., 2010), while a direct stimulation study had indicated the role of anterior superior temporal cortex in comprehension but not auditory perception (Matsumoto et al., 2011).

We found that presentation of FAS was associated with greater activity in the articulatory-phonological network, encompassing bilateral inferior frontal gyri, insula, and the right motor cortex. The inferior frontal gyrus is thought to be responsible for mapping auditory signals onto articulatory gestures (Myers et al., 2009; Lee et al., 2012; Chevillet et al., 2013). It has been suggested that the role of the inferior frontal gyrus is defined by the linkage between motor observation and imitation, which allows for abstraction of articulatory gestures from the auditory signals, along with the motor cortex and the insula (Ackermann and Riecker, 2004, 2010; Molnar-Szakacs et al., 2005; Pulvermüller, 2005; Pulvermüller et al., 2005, 2006; Skipper et al., 2005; Galantucci et al., 2006; Meister et al., 2007; Iacoboni, 2008; Kilner et al., 2009; Pulvermüller and Fadiga, 2010). On the other hand, both fMRI and transcranial magnetic stimulation (TMS) studies have indicated a functional heterogeneity within the inferior frontal cortex, which includes semantic processing (Homae et al., 2002; Devlin et al., 2003; Gough et al., 2005). The FAS and NAS stimuli had been controlled for syntactic, semantic, and phonological complexity (Calandruccio and Smiljanic, 2012). Since the task for each stimulus had also been identical (clarity rating), the increased activation across the speech processing network—including the superior temporal cortex and the articulatory network—during FAS perception is interpreted to reflect decreased neural efficiency for FAS processing (Grabner et al., 2006b; Rypma et al., 2006; Neubauer and Fink, 2009).

IMPLICIT ASIAN-FOREIGN ASSOCIATION ASSOCIATED WITH EARLY SPECTROTEMPORAL ANALYSIS

Previous behavioral studies have shown that FAS perception is modulated not only by the signal-driven factors, but also by the listener-driven factors. These listener factors can include listeners' familiarity and experience with the talkers (Bradlow and Pisoni, 1999) or language experience (Bradlow and Pisoni, 1999; Bent and Bradlow, 2003). Multiple studies have shown that listeners are also sensitive to the information regarding talker properties (Campbell-Kibler, 2010; Drager, 2010), either through explicit labels (Niedzielski, 1999; Hay et al., 2006a; Floccia et al., 2009) or facial cues (Strand, 1999; Hay et al., 2006b; Drager, 2011; McGowan, 2011; Yi et al., 2013). Listeners vary in their susceptibility to these talker cues (Hay et al., 2006b), which can override

their explicit knowledge (Hay et al., 2006a). Accordingly, listeners' implicit association between faces and foreignness (Greenwald et al., 1998, 2009; McConnell and Leibold, 2001; Bertrand et al., 2005; Devos and Banaji, 2005; Kinoshita and Peek-O'leary, 2005) modulates FAS perception only when the faces are present, through a neural mechanism hitherto unknown (Yi et al., 2013). In this study, the IAT was used to measure the degree of listener bias in which the East Asian faces are associated with foreignness of the speakers (Greenwald et al., 1998, 2009; McConnell and Leibold, 2001; Bertrand et al., 2005; Devos and Banaji, 2005; Kinoshita and Peek-O'leary, 2005). Previous fMRI studies that have used IAT as a covariate have consistently showed a pattern in which higher measures of implicit association are associated with higher activation in various neural areas for dispreferred stimuli (Richeson et al., 2003; Krendl et al., 2006; Luo et al., 2006; Suslow et al., 2010).

Examining the connection between FAS perception and listener bias, we found that listeners' implicit Asian-foreign association was reflected in the functional activity in the right primary auditory cortex. Participants with higher IAT scores showed greater activity in the primary auditory cortex for Korean-accented sentences when audiovisual information was presented. The primary auditory cortex is the site for early spectrotemporal analysis for the speech signal, sensitive to acoustic properties of the signal (Strainer et al., 1997), as well as task demands (Fritz et al., 2003, 2005), attention (Jäncke et al., 1999), context (Javitt et al., 1994), and training effects (Recanzone et al., 1993). In contrast, IAT scores were not associated with the activity in the inferior frontal gyrus. Past findings indicated that individual listeners' perceived talker properties from pictorial stimuli differentially modulate the perceptual experience (Hay et al., 2006b). In the case of FAS, the presentation of race-matched faces enhanced perception of Chinese-accented English speech (McGowan, 2011), and the individual variability in implicit Asian-foreign association predicted Korean-accented speech intelligibility (Yi et al., 2013). The present findings suggest that the listeners' implicit bias for associating Asian speakers with foreignness may be related to the early neural processing for FAS, specifically low-level spectrotemporal analysis of the acoustic properties of the signal.

LIMITATIONS AND FUTURE DIRECTIONS

In this study, Korean-accented speech was used as the proxy for FAS. Accordingly, all foreign-accented talkers appeared East Asian. In order to extend our results to the general phenomenon of FAS perception, we propose a multifactorial design in future studies where, in addition to the stimuli produced by Caucasian native speakers and Asian nonnative speakers, those by Asian native speakers and Caucasian nonnative speakers are incorporated into the study design. Also, additional explicit questionnaire on listener experience and exposure to foreign-accented stimuli could be collected to augment our understanding of the complex nature in which underlying listener biases modulate speech perception. Finally, we acknowledge that the current study did not incorporate parametric variations on the intelligibility or accent-ness of the FAS stimuli. Therefore, it is impossible to determine whether the increased BOLD response in the speech processing areas and the anterior cingulate cortex reflects increased difficulty

in comprehension or the degree of perceived foreign accent *per se* (Pelle et al., 2004; Wong et al., 2008).

CONCLUSIONS

In this study, we presented evidence of increased computational demand for FAS perception. Changes in the reduced neural efficiency for FAS processing was associated with the variability in the underlying listener biases (Yi et al., 2013). These results suggest that implicit racial association is associated with early neural response to FAS. Future studies on speech perception should examine the contribution of visual cues and listener implicit biases in order to obtain a more comprehensive understanding of the phenomenon of FAS processing.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute On Deafness And Other Communication Disorders of the National Institutes of Health under Award Number R01DC013315 awarded to Bharath Chandrasekaran as well as the Longhorn Innovation Fund for Technology to Bharath Chandrasekaran and Rajka Smiljanic. The authors thank Kirsten Smayda, Jasmine E. B. Phelps, and Rachael Gilbert for significant contributions in data collection and processing; the faculty and the staff of the Imaging Research Center at the University of Texas at Austin for technical support and counsel; the Texas Advanced Computing Center at The University of Texas at Austin for providing computing resources.

REFERENCES

- Abrams, D. A., Ryali, S., Chen, T., Balaban, E., Levitin, D. J., and Menon, V. (2013). Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's Areas. *Cereb. Cortex* 23, 1703–1714. doi: 10.1093/cercor/bhs165
- Ackermann, H., and Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain Lang.* 89, 320–328. doi: 10.1016/S0093-934X(03)00347-X
- Ackermann, H., and Riecker, A. (2010). The contribution (s) of the insula to speech production: a review of the clinical and functional imaging literature. *Brain Struct. Funct.* 214, 419–433. doi: 10.1007/s00429-010-0257-x
- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. *Brain Lang.* 122, 42–54. doi: 10.1016/j.bandl.2012.04.014
- Adank, P., Davis, M. H., and Hagoort, P. (2012a). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia* 50, 77–84. doi: 10.1016/j.neuropsychologia.2011.10.024
- Adank, P., Noordzij, M. L., and Hagoort, P. (2012b). The role of planum temporale in processing accent variation in spoken language comprehension. *Hum. Brain Mapp.* 33, 360–372. doi: 10.1002/hbm.21218
- Adank, P., and Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: adapting to time-compressed speech. *Neuroimage* 49, 1124–1132. doi: 10.1016/j.neuroimage.2009.07.032
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.
- Anderson-Hsieh, J., Johnson, R., and Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Lang. Learn.* 42, 529–555. doi: 10.1111/j.1467-1770.1992.tb01043.x
- Anderson-Hsieh, J., and Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension*. *Lang. Learn.* 38, 561–613. doi: 10.1111/j.1467-1770.1988.tb00167.x
- Anderson, M. J., and Robinson, J. (2001). Permutation tests for linear models. *Aust. N.Z. J. Stat.* 43, 75–88. doi: 10.1111/1467-842X.00156

- Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *J. Acoust. Soc. Am.* 133, EL174–EL180. doi: 10.1121/1.4789864
- Bates, D., Maechler, M., and Bolker, B. (2012). *lme4: Linear Mixed-Effects Models Using Eigen and Eigen*. Available online at: <http://CRAN.R-project.org/package=lme4>
- Bent, T., and Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am.* 114, 1600–1610. doi: 10.1121/1.1603234
- Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *Am. Econ. Rev.* 94–98. doi: 10.1257/000282805774670365
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. doi: 10.1016/j.cognition.2007.04.005
- Bradlow, A. R., and Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *J. Acoust. Soc. Am.* 106, 2074–2085. doi: 10.1121/1.427952
- Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging* 18, 32–42. doi: 10.1109/42.750253
- Calandruccio, L., and Smiljanic, R. (2012). New sentence recognition materials developed using a basic non-native english lexicon. *J. Speech Lang. Hear. Res.* 55, 1342–1355. doi: 10.1044/1092-4388(2012/11-0260)
- Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Lang. Linguist. Compass* 4, 377–389. doi: 10.1111/j.1749-818X.2010.00201.x
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432. doi: 10.1038/nn.2641
- Chevillet, M. A., Jiang, X., Rauschecker, J. P., and Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *J. Neurosci.* 33, 5208–5215. doi: 10.1523/JNEUROSCI.1870-12.2013
- Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131
- Devlin, J. T., Matthews, P. M., and Rushworth, M. F. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 15, 71–84. doi: 10.1162/089992903321107837
- Devos, T., and Banaji, M. R. (2005). American = white? *J. Pers. Soc. Psychol.* 88:447. doi: 10.1037/0022-3514.88.3.447
- Drager, K. (2010). Sociophonetic variation in speech perception. *Lang. Linguist. Compass* 4, 473–480. doi: 10.1111/j.1749-818X.2010.00210.x
- Drager, K. (2011). Speaker age and vowel perception. *Lang. Speech* 54, 99–121. doi: 10.1177/0023830910388017
- Eickhoff, S. B., Heim, S., Zilles, K., and Amunts, K. (2006). Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage* 32, 570–582. doi: 10.1016/j.neuroimage.2006.04.204
- Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., et al. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* 36, 511–521. doi: 10.1016/j.neuroimage.2007.03.060
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034
- Floccia, C., Butler, J., Goslin, J., and Ellis, L. (2009). Regional and foreign accent processing in English: can listeners adapt? *J. Psycholinguist. Res.* 38, 379–412. doi: 10.1007/s10936-008-9097-8
- Freedman, D., and Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *J. Bus. Econ. Stat.* 1, 292–298.
- Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., and Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex* 13, 170–177. doi: 10.1093/cercor/13.2.170
- Fritz, J., Elhilali, M., and Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.* 206, 159–176. doi: 10.1016/j.heares.2005.01.015
- Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141
- Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377. doi: 10.3758/BF03193857
- Goh, J. O., Suzuki, A., and Park, D. C. (2010). Reduced neural selectivity increases fMRI adaptation with age during face discrimination. *Neuroimage* 51, 336–344. doi: 10.1016/j.neuroimage.2010.01.107
- Golestani, N., and Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* 21, 494–506. doi: 10.1016/j.neuroimage.2003.09.071
- Gough, P. M., Nobre, A. C., and Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci.* 25, 8010–8016. doi: 10.1523/JNEUROSCI.2307-05.2005
- Grabner, G., Janke, A. L., Budge, M. M., Smith, D., Pruessner, J., and Collins, D. L. (2006a). “Symmetric atlas and model based segmentation: an application to the hippocampus in older adults,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006*, eds R. Larsen, M. Nielsen, and J. Sporring (Berlin: Springer), 58–66. doi: 10.1007/11866763_8
- Grabner, R. H., Neubauer, A. C., and Stern, E. (2006b). Superior performance and neural efficiency: the impact of intelligence and expertise. *Brain Res. Bull.* 69, 422–439. doi: 10.1016/j.brainresbull.2006.02.009
- Grant, K. W., and Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 108, 1197–1208. doi: 10.1121/1.1288668
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74:1464. doi: 10.1037/0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85:197. doi: 10.1037/0022-3514.85.2.197
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17. doi: 10.1037/a0015575
- Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., and Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16494–16499. doi: 10.1073/pnas.1008121107
- Hay, J., Nolan, A., and Drager, K. (2006a). From fush to feesh: exemplar priming in speech perception. *Linguist. Rev.* 23, 351–379. doi: 10.1515/TLR.2006.014
- Hay, J., Warren, P., and Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *J. Phon.* 34, 458–484. doi: 10.1016/j.wocn.2005.10.001
- Hayasaka, S., and Nichols, T. E. (2003). Validating cluster size inference: random field and permutation methods. *Neuroimage* 20, 2343–2356. doi: 10.1016/j.neuroimage.2003.08.003
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Homae, F., Hashimoto, R., Nakajima, K., Miyashita, Y., and Sakai, K. L. (2002). From perception to sentence comprehension: the convergence of auditory and visual information of language in the left inferior frontal cortex. *Neuroimage* 16, 883–900. doi: 10.1006/nimg.2002.1138
- Iacoboni, M. (2008). The role of premotor cortex in speech perception: evidence from fMRI and rTMS. *J. Physiol. Paris* 102, 31–34. doi: 10.1016/j.jphysparis.2008.03.003
- Jäncke, L., Mirzazade, S., and Joni Shah, N. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci. Lett.* 266, 125–128. doi: 10.1016/S0304-3940(99)00288-8
- Javit, D. C., Steinschneider, M., Schroeder, C. E., Vaughan, H. G. Jr., and Arezzo, J. C. (1994). Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. *Brain Res.* 667, 192–200. doi: 10.1016/0006-8993(94)91496-6
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Jenkinson, M., Pechaud, M., and Smith, S. (2005). “BET2: MR-based estimation of brain, skull and scalp surfaces,” in *Eleventh Annual Meeting of the Organization for Human Brain Mapping* (Toronto, ON).
- Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156. doi: 10.1016/S1361-8415(01)00036-6

- Kennedy, F. E. (1995). Randomization tests in econometrics. *J. Bus. Econ. Stat.* 13, 85–94.
- Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., and Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *J. Neurosci.* 29, 10153–10159. doi: 10.1523/JNEUROSCI.2668-09.2009
- Kinoshita, S., and Peek-O'leary, M. (2005). Does the compatibility effect in the race Implicit Association Test reflect familiarity or affect? *Psychon. Bull. Rev.* 12, 442–452. doi: 10.3758/BF03193786
- Krendl, A. C., Macrae, C. N., Kelley, W. M., Fugelsang, J. A., and Heatherton, T. F. (2006). The good, the bad, and the ugly: An fMRI investigation of the functional anatomic correlates of stigma. *Soc. Neurosci.* 1, 5–15. doi: 10.1080/17470910600670579
- Lee, Y.-S., Turkeltaub, P., Granger, R., and Raizada, R. D. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J. Neurosci.* 32, 3942–3948. doi: 10.1523/JNEUROSCI.3814-11.2012
- Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., and Blair, R. J. R. (2006). The neural basis of implicit moral attitude—an IAT study using event-related fMRI. *Neuroimage* 30, 1449–1457. doi: 10.1016/j.neuroimage.2005.11.005
- Macleod, A., and Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *Br. J. Audiol.* 21, 131–141. doi: 10.3109/03005368709077786
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Matsumoto, R., Imamura, H., Inouchi, M., Nakagawa, T., Yokoyama, Y., Matsuhashi, M., et al. (2011). Left anterior temporal cortex actively engages in speech perception: a direct cortical stimulation study. *Neuropsychologia* 49, 1350–1354. doi: 10.1016/j.neuropsychologia.2011.01.023
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Lang. Cogn. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006
- McConnell, A. R., and Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *J. Exp. Soc. Psychol.* 37, 435–442. doi: 10.1006/jesp.2000.1470
- McGowan, K. B. (2011). *The Role of Socioindexical Expectation in Speech Perception*. Dissertation, University of Michigan.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696. doi: 10.1016/j.cub.2007.08.064
- Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* 36, 630–633. doi: 10.3758/BF03206543
- Molnar-Szakacs, I., Iacoboni, M., Koski, L., and Mazziotta, J. C. (2005). Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cereb. Cortex* 15, 986–994. doi: 10.1093/cercor/bhh199
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., and Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13, 684–701. doi: 10.1006/nimg.2000.0715
- Möttönen, R., Krause, C. M., Tiippana, K., and Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Res. Cogn. Brain Res.* 13, 417–425. doi: 10.1016/S0926-6410(02)00053-8
- Mumford, J. (2007). *A Guide to Calculating Percent Change With Featquery*. Unpublished Tech Report. Available online at: http://mumford.bol.ucla.edu/perchange_guide.pdf
- Munro, M. J., and Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x
- Munro, M. J., and Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Lang. Speech* 38, 289–306.
- Munro, M. J., and Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Stud. Second Lang. Acq.* 23, 451–468.
- Myers, E. B., Blumstein, S. E., Walsh, E., and Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol. Sci.* 20, 895–903. doi: 10.1111/j.1467-9280.2009.02380.x
- Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex* 13, 1362–1368. doi: 10.1093/cercor/bhg083
- Neubauer, A. C., and Fink, A. (2009). Intelligence and neural efficiency. *Neurosci. Biobehav. Rev.* 33, 1004–1023. doi: 10.1016/j.neubiorev.2009.04.001
- Nichols, T. E., and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25. doi: 10.1002/hbm.1058
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psychol.* 18, 62–85. doi: 10.1177/0261927X99018001005
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex* 20, 2486–2495. doi: 10.1093/cercor/bhp318
- Peelle, J. E., McMillan, C., Moore, P., Grossman, M., and Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI. *Brain Lang.* 91, 315–325. doi: 10.1016/j.bandl.2004.05.007
- Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., et al. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 16, 125–128. doi: 10.1097/00001756-200502080-00010
- Powell, M. J. (2009). *The BOBYQA Algorithm for Bound Constrained Optimization Without Derivatives*. Cambridge NA Report NA2009/06. Cambridge: University of Cambridge.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706
- Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811
- Pulvermüller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797. doi: 10.1111/j.1460-9568.2005.03900.x
- Pulvermüller, F., Huss, M., Kherif, F., Del Prado Martin, F. M., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7865–7870. doi: 10.1073/pnas.0509989103
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- Recanzone, G. A., Schreiner, C., and Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13, 87–103.
- Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nat. Neurosci.* 6, 1323–1328. doi: 10.1038/nn1156
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153. doi: 10.1093/cercor/bhl024
- Rypma, B., Berger, J. S., Prabhakaran, V., Martin Bly, B., Kimberg, D. Y., Biswal, B. B., et al. (2006). Neural correlates of cognitive efficiency. *Neuroimage* 33, 969–979. doi: 10.1016/j.neuroimage.2006.05.065
- Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., et al. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci. Lett.* 127, 141–145. doi: 10.1016/0304-3940(91)90914-F
- Schmid, P. M., and Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *J. Speech Lang. Hear. Res.* 42, 56–64. doi: 10.1044/jslhr.4201.56
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400
- Sidas, S. K., Alexander, J. E., and Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *J. Acoust. Soc. Am.* 125, 3306–3316. doi: 10.1121/1.3101452
- Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* 25, 76–89. doi: 10.1016/j.neuroimage.2004.11.006

- Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051
- Strainer, J. C., Ulmer, J. L., Yetkin, F. Z., Haughton, V. M., Daniels, D. L., and Millen, S. J. (1997). Functional MR of the primary auditory cortex: an analysis of pure tone activation and tone discrimination. *Am. J. Neuroradiol.* 18, 601–610.
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *J. Lang. Soc. Psychol.* 18, 86–100. doi: 10.1177/0261927X99018001006
- Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- Suslow, T., Kugel, H., Reber, H., Bauer, J., Dannlowski, U., Kersting, A., et al. (2010). Automatic brain response to facial emotion as a function of implicitly and explicitly measured extraversion. *Neuroscience* 167, 111–123. doi: 10.1016/j.neuroscience.2010.01.038
- Tajima, K., Port, R., and Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *J. Phon.* 25, 1–24. doi: 10.1006/jpho.1996.0031
- Van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1181–1186. doi: 10.1073/pnas.0408949102
- Van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Commun.* 35, 103–113. doi: 10.1016/S0167-6393(00)00098-4
- Wilson, S. M., and Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* 33, 316–325. doi: 10.1016/j.neuroimage.2006.05.032
- Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* 13, 532–540. doi: 10.1016/j.tics.2009.09.003
- Wong, P. C., Uppunda, A. K., Parrish, T. B., and Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *J. Speech Lang. Hear. Res.* 51, 1026–1041. doi: 10.1044/1092-4388(2008/075)
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45, S173–S186. doi: 10.1016/j.neuroimage.2008.10.055
- Yi, H., Phelps, J. E., Smiljanic, R., and Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *J. Acoust. Soc. Am.* 134, EL387–EL393. doi: 10.1121/1.4822320

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 May 2014; accepted: 10 September 2014; published online: 08 October 2014.

Citation: Yi H, Smiljanic R and Chandrasekaran B (2014) The neural processing of foreign-accented speech and its relationship to listener bias. *Front. Hum. Neurosci.* 8:768. doi: 10.3389/fnhum.2014.00768

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Yi, Smiljanic and Chandrasekaran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multi-talker background and semantic priming effect

Marie Dekerle^{1,2*}, Véronique Boulenger^{2,3}, Michel Hoen^{2,4} and Fanny Meunier^{1,2}

¹ Laboratoire sur le Langage, le Cerveau et la Cognition, Centre National de la Recherche Scientifique, UMR 5304, Lyon, France

² University of Lyon, Lyon, France

³ Laboratoire Dynamique Du Langage, Centre National de la Recherche Scientifique, UMR5596, Lyon, France

⁴ Centre National de la Recherche Scientifique, UMR5292, Institut National de la Santé et de la Recherche Médicale, U1028, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, Lyon, France

Edited by:

Sonja A. E. Kotz, Max Planck
Institute Leipzig, Germany

Reviewed by:

Carolyn McGettigan, Royal Holloway
University of London, UK
Frederic Marmel, Universidad de
Salamanca, Spain

*Correspondence:

Marie Dekerle, Groupe de
Recherche ALP, Laboratoire sur le
Langage, le Cerveau et la Cognition,
Institut des Sciences Cognitives, 67,
Bd Pinel, 69675 BRON CEDEX,
France
e-mail: marie.dekerle@isc.cnrs.fr

The reported studies have aimed to investigate whether informational masking in a multi-talker background relies on semantic interference between the background and target using an adapted semantic priming paradigm. In 3 experiments, participants were required to perform a lexical decision task on a target item embedded in backgrounds composed of 1–4 voices. These voices were Semantically Consistent (SC) voices (i.e., pronouncing words sharing semantic features with the target) or Semantically Inconsistent (SI) voices (i.e., pronouncing words semantically unrelated to each other and to the target). In the first experiment, backgrounds consisted of 1 or 2 SC voices. One and 2 SI voices were added in Experiments 2 and 3, respectively. The results showed a semantic priming effect only in the conditions where the number of SC voices was greater than the number of SI voices, suggesting that semantic priming depended on prime intelligibility and strategic processes. However, even if backgrounds were composed of 3 or 4 voices, reducing intelligibility, participants were able to recognize words from these backgrounds, although no semantic priming effect on the targets was observed. Overall this finding suggests that informational masking can occur at a semantic level if intelligibility is sufficient. Based on the Effortfulness Hypothesis, we also suggest that when there is an increased difficulty in extracting target signals (caused by a relatively high number of voices in the background), more cognitive resources were allocated to formal processes (i.e., acoustic and phonological), leading to a decrease in available resources for deeper semantic processing of background words, therefore preventing semantic priming from occurring.

Keywords: semantic priming, informational masking, cocktail party, cognitive load, effortfulness hypothesis

INTRODUCTION

In daily life, speech is rarely perceived in silence, but with interference from wind, music or other people's conversation. Although often used to study psychoacoustic topics (Brungart, 2001; Brungart et al., 2001; McDermott, 2009), speech-in-noise and cocktail party situations (i.e., speech-in-speech, Cherry, 1953) also appear to be interesting paradigms to tackle linguistic processes and competition occurring between backgrounds and targets (Hoen et al., 2007; Boulenger et al., 2010). Our study aimed to investigate the extent to which multi-talker background is processed semantically when listening to speech-in-speech and therefore how the cocktail party situation can be used to study the automaticity of word semantic activation.

The cocktail party situation is described as involving two types of masking effects: energetic and informational masking (Brungart, 2001). Energetic masking relies on the spectro-temporal features of sounds and results from different sounds stimulating the same part of the cochlea at the same time so that one of them cannot be heard (i.e., as two signals increasingly share spectro-temporal characteristics, energetic masking becomes more efficient). In multi-talker background situations,

the magnitude of energetic masking is proportional to the number of voices that comprise the background (Simpson and Cooke, 2005). Informational masking, however, usually refers to masking effects that cannot be attributed to energetic masking. Specifically, it is related to the overlap of information carried by the different signals at a higher level (e.g., lexical level and working memory; see Durlach, 2006; Cooke et al., 2008; Mattys et al., 2009; Mattys and Wiget, 2011). Whereas background noise mainly elicits energetic masking, a speech background produces both energetic and informational masking (Brungart et al., 2006). Despite the masking, it is still possible to detect and recognize a word or linguistic token embedded in a babble. Of course, as more voices are present in the babble, participants become less accurate (Freyman et al., 2004). However, it is interesting to note that Simpson and Cooke (2005) showed, using a—6 dB SNR, that intelligibility decreases as a monotonic function of the number of speakers in babbles of up to 8 voices. Specifically, participants' accuracy to detect the target token decreases as the number of voices increases up to 8 voices. Further increasing the number of voices does not lead to a decrease in accuracy. These results suggest that if energetic masking is too high, informational masking decreases with the

diminution of the available linguistic cues. For example, with more than 8 talkers, phonetic cues are not or less available and therefore cannot be attributed incorrectly to the target.

The first aim of this paper is to test whether semantic features are involved in informational masking. It has been established that informational masking is not monolithic and occurs at many linguistic levels. Indeed, a multi-talker background will create less interference on a target word, if it is pronounced in a different language (Van Engen and Bradlow, 2007; Brouwer et al., 2012) and different languages will not have the same masking power (Gautreau et al., 2013). By manipulating the number of talkers in the background, Boulenger et al. (2010) revealed lexical competitions between a 2-talker background and target speech using a lexical decision task. Increasing the number of voices in the background, however, led to the disappearance of lexical interference because of increased energetic masking (i.e., words from the background became less intelligible and therefore competed less with target processing). However, using the same paradigm but with an intelligibility task, Hoen et al. (2007) showed that lexical processing of a background could be performed with up to 4 concurrent voices; beyond that threshold, masking was too high and seemed to prevent linguistic processes. Although it has been shown that phonological and lexical information contribute to the informational masking effect, our experiments tested the role of semantic information.

Processing of the background's semantic content has already been highlighted with 2 talkers, pronouncing either semantically correct sentences (i.e., "rice is often served in round bowls") or incorrect sentences (i.e., "the great car met the milk," Brouwer et al., 2012). Semantic incoherence in the background impacts the recognition of the target sentence. This result suggests that the background signal with 2 talkers is processed semantically. Our experiments aimed to identify how many talkers are allowed in this semantic processing using words and how semantic information from the backgrounds interferes with the identification of target words.

The ability to semantically process auditory words presented outside of the attentional focus is traditionally studied using dichotic listening. This paradigm allows to study pure informational masking as no energetic masking occurs in dichotic listening. However, discrepant results have been reported (Cherry, 1953; Lewis, 1970; Eich, 1984; Wood et al., 1997; Dupoux et al., 2003). In 1984, Eich showed a semantic effect on the recognition of words presented in the unattended channel. However, this effect resulted, at least partially, from an attentional shift toward the to-be-ignored channel (as suggested by Wood et al., 1997). As the speaker rate was very slow in Eich's experiment (85 words per minute), it allowed participants to listen to the supposedly unattended channel without disturbing the primary task (in the case of Eich's study, a shadowing task). In replicating Eich's study using the same speech rate, Wood et al. (1997) observed the same semantic effect; however, it disappeared if the speaker's rate was increased to 170 words per minute, corresponding to a more ecologically valid rate. The authors concluded that as this faster speech rate demanded more cognitive resources, participants could no longer shift attention to the unattended channel while performing the primary task, suggesting that at least in dichotic

listening, informational masking does not involve semantic information. The issue raised by this paradigm is that the spatial separation of auditory signals creates a masking release compared to a binaural condition and therefore facilitates stream segregation that could prevent competition between the to-be-ignored and target speech (Drullman and Bronkhorst, 2000; Hawley et al., 2004).

Concerning semantic activation and according to traditional theoretical models, semantic memory is organized into networks. The recognition of one word leads to its activation in semantic memory, and this activation is supposed to spread automatically to other related concepts (Collins and Quillian, 1969; Collins and Loftus, 1975). This supposition is derived from semantic priming paradigms, shown in auditory and visual modalities, in which the presentation of prime word leads to faster recognition of a semantically related target word (Meyer and Schvaneveldt, 1971; Donnenwerth-Nolan et al., 1981; Radeau, 1983; Schacter and Church, 1992; Radeau et al., 1998; Spruyt et al., 2012). For example, the presentation of the prime "nurse" before the target "doctor" facilitates the recognition of the target word "doctor" compared to a condition in which the prime is unrelated to the target (Meyer and Schvaneveldt, 1971). Adapting this paradigm to the cocktail party situation will allow us to investigate if the semantic content of the background is processed and interferes with the target word, despite decreased intelligibility. Some background words will therefore act as primes.

In the current study, we used the rationale of a priming paradigm by manipulating the association between words pronounced in the background and target words. Additionally, we varied the amount of masking to evaluate how it modulates semantic priming effects. Participants were required to perform a lexical decision task on a target item (i.e., decide whether the target item is a word or a pseudo-word) embedded in backgrounds composed of 1 to 4 voices depending on the experiment. These voices could pronounce words that were semantically related to each other and that were related or unrelated to the target. They acted as primes and were called Semantically Consistent (SC) voices. Additional voices pronounced words that were always unrelated to each other and unrelated to the target, acting as maskers. They were called Semantically Inconsistent (SI) voices.

Across experiments, we manipulated the ratio between SC and SI voices. The aim was to test the preservation of the semantic processing of SC voices despite increased masking (i.e., more SI voices). In Experiment 1, backgrounds were composed of 1 or 2 SC voices. In Experiments 2 and 3, respectively, 1 and 2 SI voices were added to each background to increase masking and therefore, decrease the intelligibility of the SC voices. Consequently, in Experiment 2, backgrounds in one condition consisted of 1 SC voice and 1 SI voice and in a second condition of 2 SC voices and 1 SI voice. In Experiment 3 they comprised 1 SC voice and 2 SI voices in one condition and 2 SC voices and 2 SI voices in the other condition.

Overall increasing the number of voices allowed us to examine if and how semantic priming can be impacted by the increase in the number of talkers in the background. Additionally, the variation in the number of SC voices compared to the number of SI voices allowed us to study the effect of prime saliency on semantic

processing and therefore its participation in informational masking. Indeed, across experiments, backgrounds can consist of the same number of voices whereas the number of SC voices compared to the number of SI voices could differ (e.g., 3 voices in the background: either 2SC/1SI in Experiment 2 or 1SC/2SI in Experiment 3).

If semantic processing can occur automatically, semantic priming should be observed at least as long as background words are intelligible and should not be disturbed by increased masking and decreased prime saliency. Indeed, automaticity is defined as a strategy free processing that occurs without using the resources of a limited capacity central processor (Neely, 1977). Therefore, if semantic processing is strategy free, it should occur even if participants are not aware that a given word is presented to them (as is done in visual modality in classical masked priming paradigms, see Forster and Davis, 1984).

EXPERIMENT 1

The aim of this experiment was to first establish set up and test our paradigm and experimental materials. Backgrounds were composed of 1 or 2 SC voices that pronounced words sharing semantic features with each other. In the related condition, target words belonged to the same semantic field as the prime, but they did not in the unrelated condition. We therefore expected to observe a semantic priming effect: participants should more quickly and accurately identify target words in the related compared to the unrelated condition. The second aim of this first experiment was to test if the presence of 2 voices in the background would affect participants' performance as suggested by the psychoacoustic literature (Brungart, 2001; Brungart et al., 2001). We therefore hypothesized that target words would be answered to more slowly and less accurately in the 2SC condition compared to the 1SC condition. Finally, we examined whether the semantic priming effect was modulated by increased energetic and informational masking caused by the augmentation of the number of voices in the background.

METHOD

Participants

Twenty-seven participants (20 females) volunteered for this experiment. All were right-handed, French native speakers and reported no known hearing or language disorder. Subjects' ages ranged from 18 to 25 years old. All participants gave written informed consent and were not aware of the experiment's purpose. They were compensated for their participation. The protocol that was used in this experiment was approved by the local ethics committee (CPP Sud-Est IV, Lyon; ID RCB: 2008-A0 0708-47).

Stimuli

Forty-eight disyllabic target words ($M_{\text{lexical frequency}} = 21.94$ per million, $SD = 18.75$ according to the French database Lexique 3, New et al., 2001) were selected, and each word belonged to a specific semantic field (e.g., *CAROTTE* "carrot"; *MÉTRO* "subway"). Each target word was matched to 10 words belonging to the same semantic field (e.g., *CAROTTE* "carrot" was associated with *légume*, *chou*, *céleri*, *salade*, *tomate* "vegetable, cabbage,

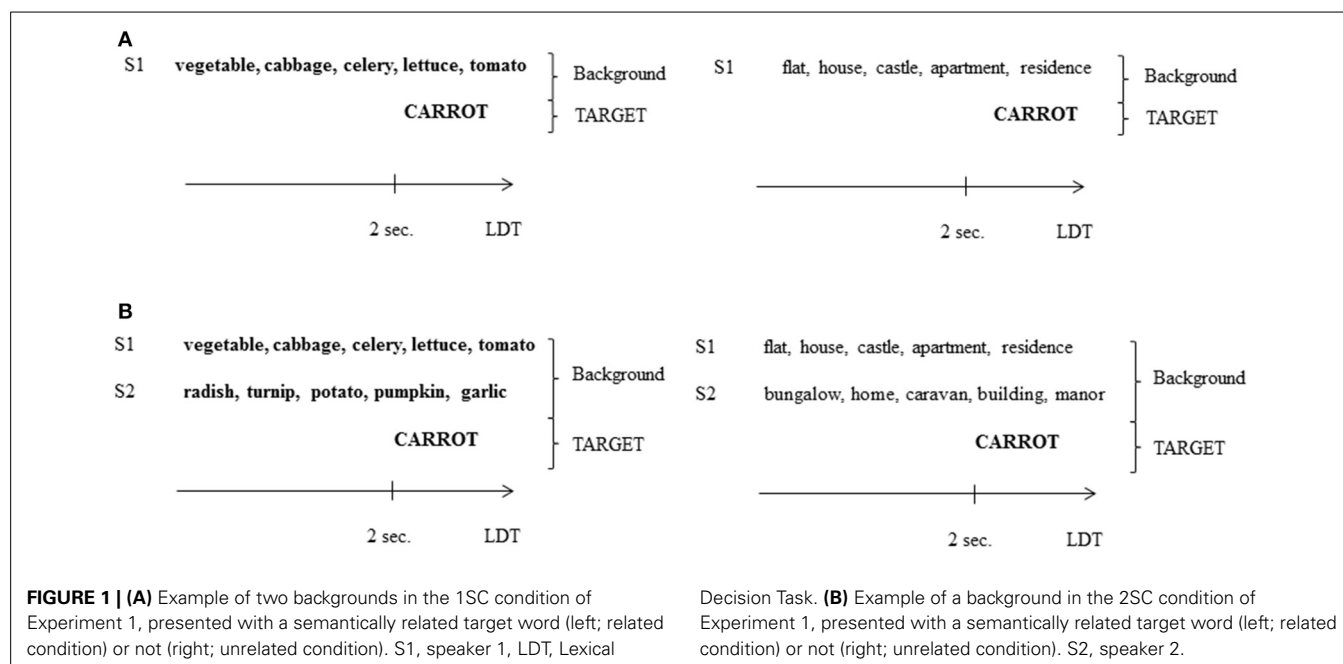
celery, lettuce, tomato"). As participants had to perform a lexical decision task, 48 pseudo-words respecting French phonotactic rules were created (e.g., *PLARO*, *HUMEL*). Ten words sharing semantic features with each other were arbitrarily associated with each pseudo-word target, resulting in a total of 96 groups of 10 words (See Supplementary Material) ($M_{\text{lexical frequency}} = 21.86$, $SD = 18.20$). As each background comprised 1 or 2 SC voices (related or not to the target), each group was divided into two subgroups of 5 words one of the subgroups was spoken by a first speaker (S1), and the other by a second speaker (S2).

Target words were presented with a semantically related (related condition) or semantically unrelated background (unrelated condition). In the unrelated condition, SC voices pronounced words that were semantically related to each other but not to the target (see Figure 1). Backgrounds comprised 1 SC voice (1SC condition) or 2 SC voices (2SC condition). The 48 target words were divided into 4 groups of 12 words, the mean frequency did not differ significantly between the groups ($F < 1$), nor did the number of phonemes [$M = 6.97$, $SD = 5.65$; $F_{(3, 44)} = 1.1$, *n.s.*] and phonological neighbors [$M = 4.75$, $SD = 0.81$; $F_{(3, 44)} = 2.2$, *n.s.*]. Each group of twelve target words was assigned to a condition (1SC related, 1SC unrelated, 2SC related, 2SC unrelated) depending on the experimental list. The same was true for pseudo-words. Four experimental lists of 96 stimuli (i.e., 48 target words and 48 target pseudo-words) were created so that each target word was presented in each condition, but only once in a list (each participant was presented with one list only).

Targets and SC voices were recorded by 3 different French native female speakers (age: 21–22) in a sound-proof room (22 kHz, mono, 16 bits). Auditory sequences of 5 words from Speakers 1 and 2 (S1 and S2) were segmented into 3 s periods. The periods were then normalized at an intensity of 60 dB-A and mixed together to create backgrounds. All audio files were synchronized at the beginning, so all voices started to speak at the same time. However, as all voices pronounced words of different lengths, they soon became desynchronized, and there was always one speaker talking in the background. Targets recorded by Target Speaker (TS; also normalized at an intensity of 60 dB-A) were inserted 2 s after the start of the backgrounds (so that each participant always had the same exposure to the background before the target speech was presented), with a 0 dB SNR (Signal/Noise Ratio; see Figure 1). Because the backgrounds, which comprised 1 or 2 voices, generated different amounts of energy, the intensity of all stimuli was varied over a ± 3 dB range in 1 dB steps to prevent participants from predicting condition depending on individual stimuli intensity.

Procedure

Participants sat in front of a computer screen and heard the stimuli binaurally through headphones at a comfortable level (mean level 65 dB-A, ranging from 62 dB-A to 68 dB-A, normalized using an artificial ear). A fixation cross was presented on the screen at the beginning of each trial and remained on the screen during stimulus presentation. Participants were asked to listen to the stimuli to decide as quickly and accurately as possible whether the target was a word or a pseudo-word, by pressing one of two



pre-specified keys. After a response was given, a string of hash marks indicated that the trial was over; participants could then press a key to start the next trial. Half of the participants gave the response to “word” with their left hand and to “pseudo-word” with their right hand. As all participants were right-handed, they might answer faster with their right hand than their left hand. To avoid this confounding effect, the other half were given the opposite instruction. A training session composed of twelve trials (different from the experimental stimuli) preceded the test session so that participants could acclimate to the stimuli and the task.

RESULTS

Two Two-Way repeated measures analyses of variance (ANOVAs) by participants (F_1) and by items (F_2) were conducted, with Response Times (RTs, in ms) and Error Rates (ERs) for target word identification as dependent variables. We included Number of Voices in the background (1 Voice, 1SC or 2 Voices, 2SC) and Semantic Link between prime and target (related or unrelated) as within-subjects factors. Three participants were excluded from analyses because of very high ERs (more than 40%). Four target words error rates greater than 50% were also excluded from Item analyses (*POIGNET*, *RIDEAU*, *RATON*, and *RACINE* “wrist, curtain, baby rat, root”). Trials with RTs below or above 2.5 standard deviations from the individual means (4.5%) and trials in which participants made mistakes (19.5%) were not included in RTs analysis. Means and Standard Deviations (SDs) of RTs and ERs are summarized in **Table 1**.

The ANOVA by participants first revealed a significant main effect of the Number of Voices: participants were faster [$F_{1(1, 23)} = 4.25, p = 0.05$] and more accurate [$F_{1(1, 23)} = 9.12, p < 0.01$] to identify targets in the 1SC condition ($M_{RT} = 1008$ ms, $SD = 157$; $M_{ER} = 11.7\%$, $SD = 10.9$) than in the 2SC condition ($M_{RT} = 1042$ ms, $SD = 161$; $M_{ER} = 20.1\%$,

Table 1 | Means and Standard Deviations (SDs) of Response Times (RTs) and Error Rates (ERs) depending on the number of voices in the background and the semantic link between prime and target in Experiment 1.

		1SC		2SC	
		Related	Unrelated	Related	Unrelated
RT (ms)	Mean	980	1035	1014	1070
	SD	170	142	169	150
ER (%)	Mean	7.08	16.33	19.5	20.77
	SD	6.82	12.34	19.04	14.05

1SC, 1 SC voice condition; 2SC, 2 SC voices condition; related, semantic link between the prime and the target; unrelated, no semantic link between the prime and the target.

$SD = 16.5$). The Item analysis, however, did not highlight an effect of the Number of Voices on RT [$F_{2(1, 43)} = 1.9, p = 0.1$], although target words were better categorized as words in the 1SC condition [$F_{2(1, 43)} = 13.43, p < 0.001$].

The main effect of Semantic Link also appeared to be significant on RTs [$F_{1(1, 23)} = 14.24, p < 0.001$; $F_{2(1, 43)} = 4.63, p < 0.05$], participants responded faster if targets shared semantic features with the prime ($M = 997$ ms, $SD = 169$) than if they did not ($M = 1053$ ms, $SD = 145$); this resulted in a 56 ms priming effect. This effect was also significant for ERs in the participant analysis [$F_{1(1, 23)} = 3.93, p = 0.05$], and there was only a trend in the item analysis [$F_{2(1, 43)} = 3.07, p < 0.10$]. Participants tended to be more accurate in the related condition ($M = 15.4\%$, $SD = 15.4$) than in the unrelated condition ($M = 18.55\%$, $SD = 13.2$). There was no significant interaction between the two factors for RTs ($F_1 < 1$; $F_2 < 1$) and ERs [$F_{1(1, 23)} = 2.34, n.s.$; $F_{2(1, 43)} = 1.97, n.s.$], suggesting that the semantic priming effect was not

modulated by the Number of Voices (one or two) in the background.

DISCUSSION

These results first highlight that participants were slowed by the increase in the number of voices in the background. This effect is certainly attributable to enhanced target masking in the two-voice condition (Brungart, 2001; Brungart et al., 2001). Interestingly, participants' performance was improved by the semantic relationship between the prime and target, and this effect was independent of the number of voices, suggesting that the increase in masking from one to two background voices, was not sufficient to prevent semantic processing. However, in this first experiment, prime was salient in both conditions (1SC voice and no SI voice or 2SC voices and no SI voice). To test whether participants could still take advantage of the semantic relationship between target and prime if the intelligibility of the SC voices was further decreased, we conducted a second experiment in which a SI voice was added to each background.

EXPERIMENT 2

This second experiment aimed to investigate whether the semantic priming effect would resist increased masking. An SI voice was therefore added to each background. This voice pronounced words sharing no semantic features with each other or with the target word, whatever the condition. The purpose was to use the same material and procedure as in Experiment 1 with the addition of mask on the SC voices. In Experiment 2, backgrounds were composed of two voices (1 SC voice + 1 SI voice) or 3 voices (2 SC voices + 1 SI voice). A deleterious effect of the number of voices on participants' performance was predicted, and we expected that the presence of SI voice would not affect the semantic priming effect if this latter effect is automatic.

Another change was made in Experiment 2 regarding target items. In Experiment 1, target items were pronounced by a female speaker, and were consequently, difficult to detect among the other female speakers (S1 and S2). These difficulties might partly explain the low accuracy and long response times to target words inserted in babbles that were only composed of one or two voices. To avoid flux segregation difficulties (Festen and Plomp, 1990; Brungart et al., 2001), target items were therefore pronounced by a male speaker (Target Speaker 2; TS2) in the two following experiments.

METHOD

Participants

Twenty-four right-handed French native speakers (18 females), aged 18–30 years, participated in this second experiment. They had no known auditory or language disorders. All participants gave their written informed consent and were compensated for their participation. None of the participants had been tested in Experiment 1 and they were not aware of the aim of the study before testing.

Stimuli

To add an SI voice to each background used in Experiment 1, 96 groups of 5 words ($M_{\text{lexical frequency}} = 18.15$, $SD = 9.75$), not

semantically related to each other, were generated. Average lexical frequency did not differ between SC voices (from Experiment 1) and the SI voice as shown by an ANOVA [$F_{(2, 190)} = 1.16$, *n.s.*]. Each group was selected to mask a specific prime (composed of 1 or 2 SC voices), which shared no semantic link (e.g., the prime *légume, chou, céleri, salade, tomate* “vegetable, cabbage, celery, lettuce, tomato” was always presented with the SI voice pronouncing *policier, intéressant, cour, affiche, étagère* “policeman, interesting, yard, poster, shelf”). This SI voice was recorded by another French native female speaker (S3, age = 20) using the same method as in Experiment 1.

Backgrounds composed of 1 SC voice (S1) in Experiment 1 were now composed of 1 SC voice and 1 SI voice (S1 + S3), corresponding to the 1SC/1SI condition, and backgrounds composed of 2 SC voices (S1 + S2) in Experiment 1 were now composed of 2 SC voices and 1 SI voice (S1 + S2 + S3), corresponding to the 2SC/1SI condition. The 4 groups of 12 target words and pseudo-words created for Experiment 1 were used. Target words were presented in each of the 4 conditions: 1SC/1SI related, 1SC/1SI unrelated, 2SC/1SI related and 2SC/1SI unrelated. The corresponding number of pseudo-words was also presented. Four experimental lists were created so that each target word was seen in each condition but only once in a list.

Recordings of S1 and S2, used in the previous experiment, were mixed with S3 following the previously established experimental lists. Targets were recorded by a French native male speaker (Target Speaker 2; age = 20) and were inserted into backgrounds 2 s after the start of the sequence (see **Figure 2**).

Procedure

The procedure was the same as in Experiment 1.

RESULTS

Similar analyses as in Experiment 1 were performed by subjects (F_1) and by items (F_2), with RTs (in ms) and ERs for target word identification as the dependent variables. We included Number of Voices in the background (two voices, 1SC/1SI or three voices, 2SC/1SI) and Semantic Link between prime and target (related or unrelated) as within-subjects factors. As in Experiment 1, target words for which less than 50% of participants answered correctly were not analyzed. The target words *BILLET* and *CHIGNON* (“ticket” and “bun”) were therefore not included. Moreover, 15% of data were excluded (13% of errors and 2% of extremes values) from RT analysis. Mean RTs and ERs with SDs are summarized in **Table 2**.

The Two-Way repeated measures ANOVAs showed a significant main effect of the Number of Voices in the backgrounds [$F_{(1, 23)} = 6.08$, $p < 0.05$; $F_{(2, 45)} = 4.34$, $p < 0.05$]. Participants responded faster to target words if backgrounds comprised 2 voices ($M = 946$ ms, $SD = 157$) compared to 3 voices ($M = 973$ ms, $SD = 175$). This main effect was also significant for ERs [$F_{(1, 23)} = 7.18$, $p < 0.05$] but only in the participants' analysis [$F_{(2, 45)} = 3.72$, $p < 0.10$]. Responses were more accurate in the 1SC/1SI condition ($M = 8.6\%$, $SD = 8.2$) than in the 2SC/1SI condition ($M = 12.5\%$, $SD = 9.1$).

The main effect of Semantic Link was also significant [$F_{(1, 23)} = 5.32$, $p < 0.05$; $F_{(2, 45)} = 5.38$, $p < 0.05$], responses

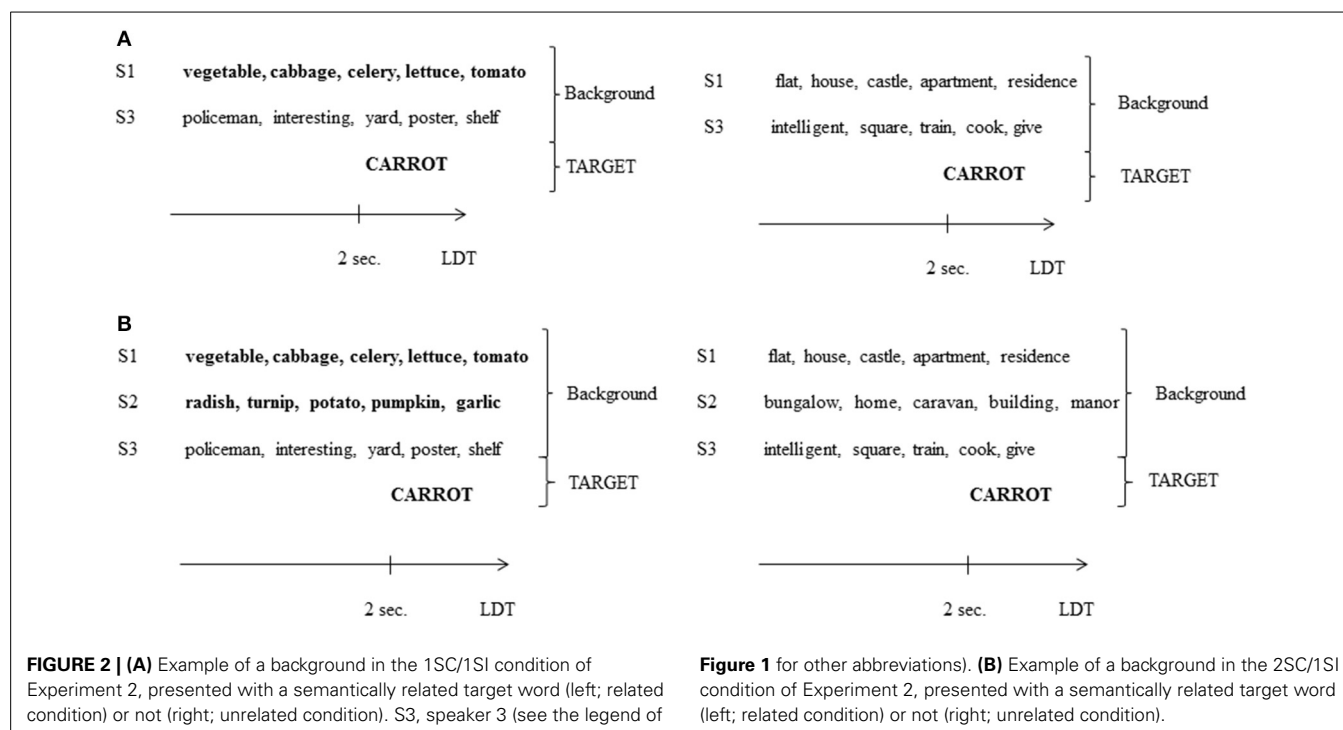


Table 2 | Means and SDs of RTs and ERs depending of the number of voices in the background and the semantic link between prime and target in Experiment 2.

		1SC/1SI		2SC/1SI	
		Related	Unrelated	Related	Unrelated
RT (ms)	Mean	933	958	945	1000
	SD	165	153	162	187
ER (%)	Mean	6.47	10.7	10.7	14.33
	SD	7.29	9.14	9.32	8.78

1SC/1SI, 1 SC voice and 1 SI voice condition; 2SC/1SI, 2 SC voices and 1 SI voice condition.

were 40 ms faster if the prime and target shared semantic features ($M = 939$ ms, $SD = 162$) compared to if they did not share features ($M = 979$ ms, $SD = 170$). This effect also reached significance for ERs in the participants' analysis [$F_{(1, 23)} = 5.33$, $p < 0.05$; $F_{(2, 45)} = 1.98$, $p = 0.10$]. ERs decreased by 3.9% if the prime and target were semantically related ($M = 8.61\%$, $SD = 8.5$ in the related condition and $M = 12.53\%$, $SD = 8.8$ in the unrelated condition).

There was no significant interaction between the Number of Voices in the backgrounds and the Semantic Link between prime and target for RTs [$F_{(1, 23)} = 1.92$, $n.s.$; $F_2 < 1$] and ERs ($F_1 < 1$; $F_2 < 1$), indicating that the priming effect was not affected by the increase in the number of voices in the backgrounds.

DISCUSSION

As in Experiment 1, performances improved if the prime and target were semantically related, although participants were slower

in condition 2SC/1SI than 1SC/1SI because of increased masking (Bronkhorst, 2000; Brungart, 2001; Brungart et al., 2001). Interestingly, there was again no indication that increased masking reduced the semantic priming effect. To further test this resistance of the semantic priming effect to prime intelligibility loss, we conducted a third experiment in which a second SI voice was added to each background to further decrease prime saliency. However, the intelligibility of the target item may decrease with the addition of a second SI voice in the background. However, as the target item is pronounced by a male voice and embedded in a female voice background, it is still quite salient compared to background voices (Brungart et al., 2001). Indeed, Brungart and collaborators showed that participants more easily recognize a target sentence if it was embedded in a background composed of voices of a different sex than if all the sentences (background and target) were pronounced by speakers of the same sex. Additionally, in our experiment, target items were presented at the same time (2 s after the beginning of the stimulus), and this regular timing helps participants to detect the target, as they know when to listen to it. Consequently, in our paradigm, the masking effect of the SI voice on the target item was quite low compared to its effect on SC voices.

EXPERIMENT 3

This last experiment was the same as Experiment 2, except that a second SI voice was added to each background (i.e., backgrounds comprised 3 or 4 voices). The same method was used as in Experiment 2. The aim was to further increase the masking of SC voices to test the resistance and automaticity of semantic processing. As the number of voices in the backgrounds increased (i.e., up to 4 voices), we wanted to make sure whether participants could still identify words from the background. Therefore, after the

main experiment, we asked participants to perform a recognition test. This post-test was designed to examine whether, in agreement with Hoen et al. (2007), words in the background were still intelligible. It aimed to clarify, in the case of significant semantic priming, if this effect resulted from preserved intelligibility of the prime words by testing lexical access or from automatic processing. In case of preserved intelligibility we cannot prove that the semantic effect results from automatic processing, it may be either automatic or strategic. However, if a priming effect is found without preserved intelligibility, semantic processing is automatic (as shown with masked priming paradigm in visual modality, see Dehaene et al., 1998; Naccache and Dehaene, 2001; Spruyt et al., 2012). Proof of non-intelligibility was consequently needed in the case of a significant semantic priming effect to provide evidence for an automatic process. Otherwise, we would not be able to detect whether automatic components are present in semantic processes. As in our previous experiments, we expected to observe a significant effect of the number of voices in the background; increasing the number of voices has shown to decrease intelligibility for up to 8 voices (Simpson and Cooke, 2005). No interaction between the number of voices and the semantic association between prime and target was expected, at least if the words composing the background were still intelligible.

METHOD

Participants

Twenty-four participants (19 females) were recruited for this experiment (age: 18–34). All were right-handed French native speakers and reported no hearing or speech disorder. They gave written informed consent and were compensated for their participation. None of them had participated in Experiments 1 or 2, and they were unaware of the experiment's purpose prior to testing.

Stimuli

To create an extra SI voice, pronounced by a fourth speaker (S4), 96 groups of 5 words, that were semantically unrelated to each other, were generated ($M_{\text{lexical frequency}} = 20.88$, $SD = 1.22$). The word mean frequency did not differ between the different voices (SC and SI voices; $F < 1$). As in Experiment 2, each group of words was matched to a prime (composed of 1 or 2 SC voices) with which it did not share any semantic features and was systematically presented with (e.g., the prime *légume*, *chou*, *céleri*, *salade*, *tomate* “vegetable, cabbage, celery, lettuce, tomato” was always presented with the masker *étui*, *liberté*, *drôle*, *global*, *sympathie* “case, freedom, funny, global, sympathy”). Therefore, with the addition of this second SI voice, backgrounds composed of 2 voices (S1 + S3) from Experiment 2 became 3-voice backgrounds (S1 + S3 + S4; 1 SC voice and 2 SI voices; 1SC/2SI condition) and 3-voice backgrounds (S1 + S2 + S3) from Experiment 2 became 4-voice backgrounds (S1 + S2 + S3 + S4; 2 SC voices and 2 SI voices; 2SC/2SI condition). The 4 groups of 12 target words and pseudo-words created in Experiment 1 were used and presented in the following conditions: 1SC/2SI related, 1SC/2SI unrelated, 2SC/2SI related, and 2SC/2SI unrelated. Four experimental lists were created so that each target word was presented in each condition but only once in a list.

The second SI voice (S4) was recorded by a French native female speaker (age = 23), using the same procedure as in previous experiments. S1, S2, and S3's auditory sequences were mixed with S4's. Targets used in Experiment 2 (recorded by TS2) were embedded in the backgrounds 2 s after their beginning start (see Figures 3, 4).

Recognition post-test

A recognition test was devised to test whether participants could recognize words previously heard during the experiment. Fifty words were presented to participants after the main experiment on a sheet of paper, 20 had been previously presented in the backgrounds, whereas 30 were new words not used as stimuli in the experiment. A list of new words was generated ($M_{\text{lexical frequency}} = 23.15$, $SD = 48.29$), and their lexical frequency did not significantly differ from that of previously heard words ($M_{\text{lexical frequency}} = 18.47$, $SD = 20.61$; $F < 1$).

Procedure

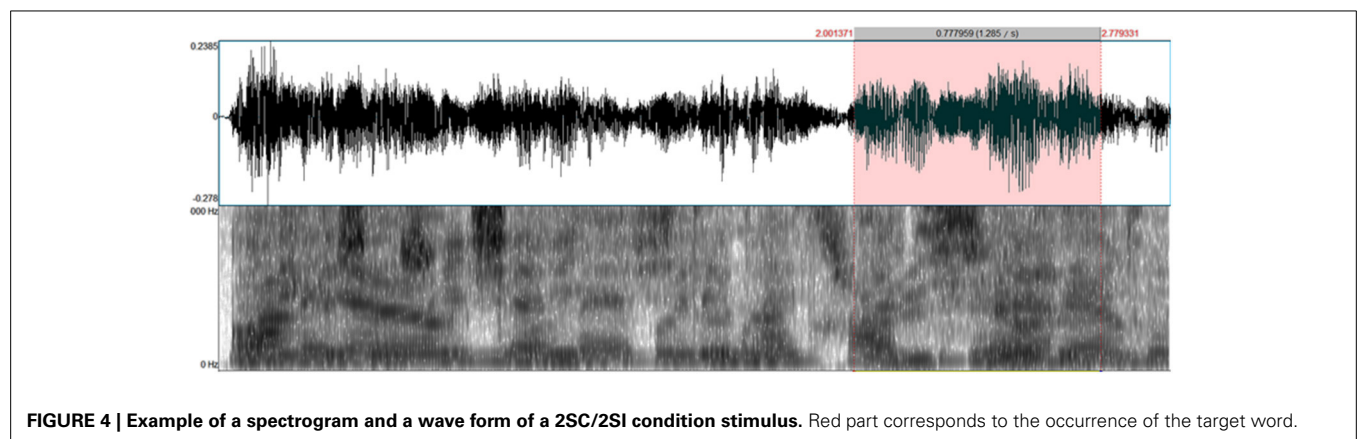
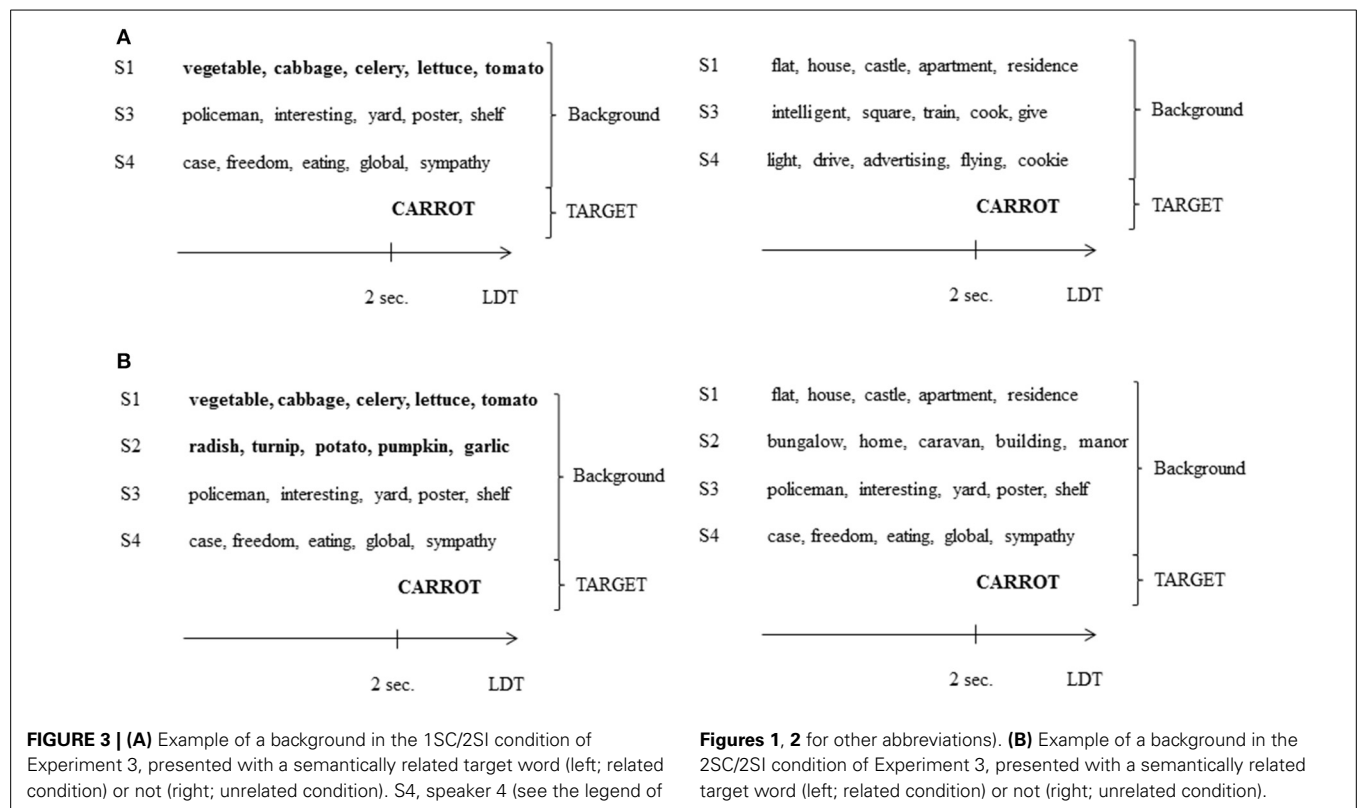
The same procedure was used as in Experiments 1 and 2. At the end of the experiment, a post-test was given to participants, instructing them to decide (i.e., write down) whether they had heard the given word during the experiment. They were asked to do the best they could but not to think about it too much, and to simply answer what they thought was right.

RESULTS

Test

The same statistical analyses as in Experiments 1 and 2 were performed by participants (F_1) and by items (F_2), with RTs (in ms) and ERs for target word identification as dependent variables. We included Number of Voices in the background (3 voices, 1SC/2SI condition or 4 voices, 2SC/2SI condition) and Semantic Link between prime and target (related or unrelated) as within-subjects factors. Target words answered correctly by less than half of participants were excluded from analyses (*CHIGNON*, *EPAULE*, and *HIBOU*, “bun, shoulder, owl”). In total, 16.8% of data were excluded from RTs analysis because of errors (15%) or extreme values (1.8%). Means and SDs for RTs and ERs are summarized in Table 3.

The analysis by participants revealed a significant main effect of the Number of Voices on RTs [$F_{1(1, 23)} = 4.01$, $p = 0.05$; $F_2 < 1$]. Participants more quickly identified target words if backgrounds were composed of 3 voices (1SC/2SI condition; $M = 925$ ms, $SD = 155$) than if they were composed of 4 voices (2SC/2SI condition; $M = 941$ ms, $SD = 155$). The ERs analysis also showed that participants responded significantly more accurately [$F_{1(1, 23)} = 6.19$, $p < 0.05$; $F_{2(1, 44)} = 5.80$, $p < 0.05$] in the 1SC/2SI condition ($M = 10.02\%$, $SD = 10.10$) than in the 2SC/2SI condition ($M = 13.21\%$, $SD = 8.95$). The main effect of Semantic Link was not significant for either RTs [$F_{1(1, 23)} = 1.51$, $n.s.$; $F_{2(1, 44)} = 3.21$, $n.s.$] or ERs [$F_{1(1, 23)} = 3.19$, $n.s.$; $F_{2(1, 44)} = 2.92$, $n.s.$]. No significant interaction emerged between the 2 factors for RTs [$F_{1(1, 23)} = 1.39$, $n.s.$; $F_{2(1, 44)} = 1.10$, $n.s.$] and ERs ($F_1 < 1$; $F_2 < 1$). These last results indicate that participants performed better in the lexical decision task, both in terms of speed and accuracy, if the backgrounds were composed of 3



voices rather than 4 voices. This finding highlights the impact of masking on target recognition. However, no significant semantic priming effect was observed in these conditions, suggesting that informational masking from 1SC/2SI and 2SC/2SI backgrounds was so efficient that it prevented semantic processing of the prime, which could therefore not affect target word identification. Additionally, in the 2SC/2SI condition, energetic masking was more important; this factor might also explain the important masking of prime and explain the lack of semantic priming.

Post-test

Analysis of the participants' answers showed a mean ER of 44.5% ($SD = 7$) and $d' = 0.26$ ($SD = 0.5$). Although these results are close to chance, both were significant results, as shown by a

one-sample t -test ($ER\ p < 0.01$; $d'\ p < 0.05$). No significant correlation was found between d' and priming effect ($r = -0.39$, $n.s.$). This finding implies that participants heard some background words, but those words were not used to improve performances. Additionally, a repeated measure ANOVA was performed with ERs as the dependent variable and number of Voices (3 or 4) in the background and Semantic Link between presented word and target as a within subjects factor. Neither the effect of the Number of Voices nor the effect of Semantic Link were significant ($F_{\text{Number of Voices}} < 1$; $F_{\text{Semantic Link}} < 1$).

DISCUSSION

In this third experiment, target word identification was again disturbed by the increase in the number of voices in the

backgrounds, confirming that masking is more efficient in the 4-voice than in the 3-voice condition. Disappearance of the semantic priming effect also suggests that the number of SC voices compared to the number of SI voices was too small. To compare the data of the 3 experiments, we considered the ratio of SC voices over the total number of voices. Therefore, the ratio of SC/total voices is 1 in 1SC and 2SC conditions; ratio 2/3 in 2SC/1SI; ratio 1/2 in 1SC/2SI and 2SC/2SI; and ratio 1/3 in 1SC/2SI. An ANCOVA on all data using the number of voices as the independent variable and the ratio of SC voices/total voices as covariate on priming effect confirmed this hypothesis (effect of ratio: $p < 0.05$): when the ratio was too low, SC voices were not salient enough for participants to perform semantic processing. In a post-test, participants were asked to perform a recognition task immediately after the experiment. Participants scored significantly better than chance at the recognition test, showing that at least some words in the background were identified. This finding is consistent with the results by Hoen et al. (2007) who showed

that in a transcription task, with up to 4 voices in the background, participants gave words from the background as responses instead of target words. Overall Experiment 3 showed that participants were unable to process the prime at a semantic level (i.e., no priming effect was observed) although the post-test results suggest that they hear it sufficiently to recognize it in a recognition post-test. This finding suggests that a word can be heard and implicitly encoded without being sufficiently deeply processed to elicit semantic priming.

POST-HOC ANALYSES

To better analyze the impact of the SC/total voices ratio, we conducted *post-hoc* analyses. A HSD Tukey test showed that up to 4 voices in the background if the ratio of SC/total voices was inferior to 1/2, there was no significant semantic priming effect (Experiment 2. 1SC/1SI; Experiment 3. 1SC/2SI and 2SC/2SI). However, if the ratio was superior to 1/2, semantic priming was significant (Experiment 1: 1SC, $p < 0.05$ Cohen's $d = 0.31$; 2SC, $p < 0.05$, Cohen's $d = 0.35$; Experiment 2. 2SC/1SI, $p < 0.05$, Cohen's $d = 0.31$; cf Figure 5 and Table 4).

GENERAL DISCUSSION

SUMMARY

The aim of this study was to investigate to what extent the semantic content of the multi-talker babble could interfere with the processing of targets in a cocktail party situation. In the three reported experiments, participants were required to perform a lexical decision task on a target item embedded in a multi-talker background. The backgrounds were composed of 1 or 2 SC voices pronouncing words that shared semantic features with each other and could be semantically related to the target or not. The ratio of SC voices over the total number of voices in the background was varied across experiments. In Experiment 1, ratios were 1/1 and 2/2 (i.e., only SC voices were presented in the background).

Table 3 | Means and SDs of RTs and ERs depending on the number of voices in the background and the semantic link between prime and target in Experiment 3.

		1SC/2SI		2SC/2SI	
		Related	Unrelated	Related	Unrelated
RT (ms)	Mean	924	928	928	953
	SD	157	158	156	157
ER (%)	Mean	7.03	11.1	13.00	15.27
	SD	7.30	9.64	11.69	7.88

1SC/2SI, 1 SC voice and 2 SI voices condition; 2SC/2SI, 2 SC voices and 2 SI voices condition.

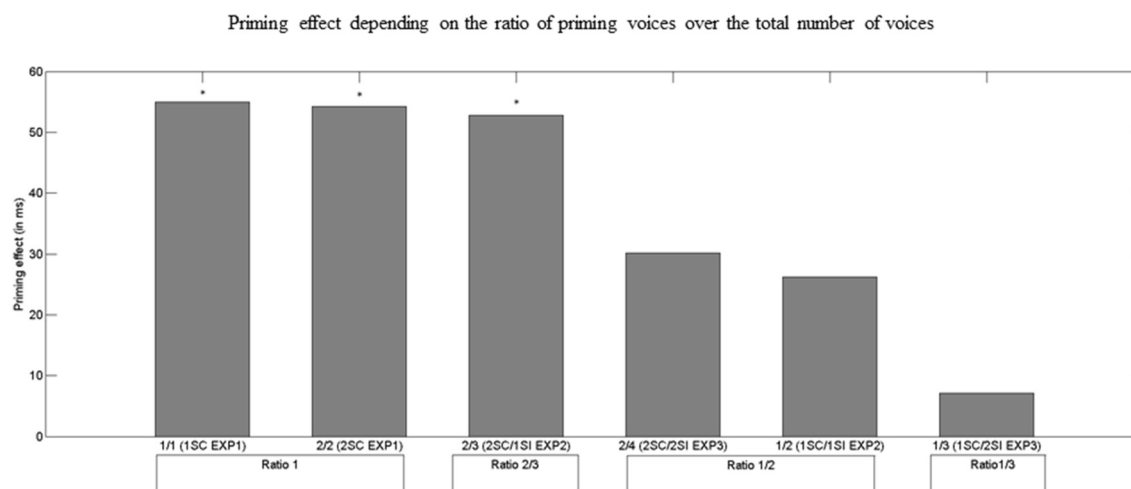


FIGURE 5 | Semantic priming effect (i.e., difference between RTs for unrelated and related targets; in ms) depending on the ratio of SC voices in the background over the total number of voices. 1SC EXP1, 1SC condition in Experiment 1 (1 SC voice; no SI voice); 2SC EXP1, 2SC condition in Experiment 1 (2 SC voices; no SI voice);

2SC/1SI EXP2, 2SC/1SI condition in Experiment 2 (2 SC voices; 1 SI voice); 1SC/1SI EXP2, 1SC/1SI condition in Experiment 2 (1 SC voice; 1 SI voice); and 2SC/2SI EXP3, 2SC/2SI condition in Experiment 3 (2 SC voices; 2 SI voices); 1SC/2SI EXP3, 1SC/2SI condition in Experiment 3 (1 SC voice; 2 SI voices).

Table 4 | Mean priming effect (in ms), SD, and *p*-value for each ratio.

Ratio	1		2/3	1/2		1/3
Condition	1SC	2SC	2SC/1SI	1SC/1SI	2SC/2SI	1SC/2SI
	(EXP1)	(EXP1)	(EXP2)	(EXP2)	(EXP3)	(EXP3)
Mean	55	56	55	25	26	4
SD	70	119	110	88	83	65
<i>P</i>	0.03*	0.03*	0.007*	0.88	0.21	0.99

* Indicates significant priming. The ratio is expressed in terms of the number of SC voices over the total number of voices in the background: 1, 1 SC voice or 2 SC voices; 2/3, 2 SC voices, 1 SI voice; 1/2, 1 SC voice, 1 SI voice or 2 SC voices, 2 SI voices; 1/3, 1 SC voice, 2 SI voices.

In Experiments 2 and 3, 1 and 2 SI voices, respectively, that pronounced semantically unrelated words, were added to backgrounds to decrease the intelligibility of the SC voices, which acted as the prime. The ratio of SC/total voices in Experiment 2 was therefore 1/2 and 2/3 and in Experiment 3 it was 1/3 and 2/4.

The main effect of the Number of Voices was significant in each experiment; participants responded faster and more accurately to target words if a smaller number of voices composed the backgrounds. This delayed response time with increasing number of voices is a well-known phenomenon (Brungart et al., 2001; Boulenger et al., 2010). As the number of voices increases, energetic masking is enhanced so that the signal is saturated, and target items are consequently more difficult to process. In addition to physical masking, more information is perceived and must be processed (i.e., informational masking), leading to slower response times.

Overall the main effect of Semantic Link between prime and target seems to be an all-or-none phenomenon. The semantic priming effect could have decreased with an increasing number of voices in the background, but our results suggest this effect does not occur, as no interaction between the Number of Voices and the Semantic Link was observed in any of the three experiments. The main effect of Semantic Link depends on the ratio of SC voices over the total number of voices in the background. A semantic priming effect emerged in our experiments spanning from 1 to 4 voices in the background but only if the number of SC voices was higher than the number of SI voices. This ratio of SC voices is interesting because it highlights the necessity of prime saliency for semantic processing to occur; it also gives an objective measure of this prime saliency across experiments. This finding suggests that informational masking does occur at the semantic level but only if the prime is sufficiently salient. If informational and energetic masking had the same role, a significant semantic priming effect would appear in conditions containing 3 voices (i.e., Experiment 2 2SC/1SI and Experiment 3 1SC/2SI). However, this effect did not occur; we therefore conclude that informational masking can be semantic only if it is sufficiently salient to be used to increase performance.

LEXICAL PROCESSING WITHOUT SEMANTIC ACTIVATION

Our results suggest that semantic processing in cocktail party situation is not automatic. In fact, it seems that in challenging

listening situations, one can hear and activate a mental representation of a word without deeply processing it at a semantic level. Background words can be semantically processed with up to 3 voices (2SC/1SI, Experiment 2); however no priming effect emerged in Experiment 3, despite the recognition post-test's results suggesting that primes were heard and recognized. This finding is consistent with the way the language system has been modeled; most models of word processing, both in the auditory and the visual modalities, suggest a distinction between lexical and semantic levels of processing (Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Grainger and Holcomb, 2009). If one considers that these processes are independent stages of word recognition, it seems reasonable that one of these stages can be reached (i.e., lexical) without strongly activating the deepest stage (i.e., semantic).

Many studies have highlighted the automaticity of semantic processing using masked semantic priming (Naccache and Dehaene, 2001; Klauer et al., 2007; Kouider and Dehaene, 2009; Spruyt et al., 2012). However, masked semantic priming is only found in very specific conditions in the visual modality and with semantic categorization tasks (see Van den Bussche et al., 2009, for a review). One can therefore assume that semantic activation remains superficial in masked priming paradigm, and only activates very close concepts such as superordinates, which is enough to create a priming effect in semantic categorization tasks. In our experiments however, a deeper semantic processing was necessary. For example, in the semantic field of birds, the SC voice pronounced: *corbeau*, *rossignol*, *cage*, *voler*, *nid* ("craw, nightingale, cage, fly, nest") and *PIGEON* ("pigeon") was the target word. In a condition of high intelligibility, these words primed the target word; however, with decreased intelligibility, if the participants only heard the word "cage," this word may have activated its superordinate (e.g., "object") but not its associates such as "bird." Consequently, it may not have primed "pigeon." The absence of a semantic priming effect in addition to the decrease in intelligibility seems to show that greater difficulty in processing auditory signals at a superficial level, causes decreased processing at a (higher) linguistic level.

Consistent with our results, and using an auditory masked priming paradigm, Kouider and Dupoux (2005) demonstrated lexical access without semantic priming. In their experiments, prime was a time-compressed word embedded in a masker composed of time reversed and compressed words. They manipulated the compression rate of the prime to vary its intelligibility. If the prime was not intelligible, they found a significant repetition priming effect on words but not non-words suggesting that this effect "involved a lexical activation of abstract word form." In the same condition of prime compression, they did not find any semantic priming effect. This finding would therefore suggest a dissociation between lexical and semantic processing.

COGNITIVE LOAD

Overall our results are compatible with the idea of cognitive load, in which simultaneously processed information and interactions can either under-load or overload the finite amount of processing capacities. As shown by our results, it is more difficult to process items with more voices in the background. This finding

might be partly due to the high perceptual load and because the target item was embedded in backgrounds that could not be completely ignored (Lavie, 1995, 2005). Processing words in the background might have been particularly cognitively effortful and, semantic processing of a word can be delayed and even prevented if participants have to perform an additional task (i.e., high cognitive load, Hohlfeld et al., 2004; Hohlfeld and Sommer, 2005; Van Petten, 2014). In Experiment 3 we argue that some background words were heard but not deeply semantically processed because it was both very demanding and irrelevant to perform the task. In Experiment 2, 2SC/1SI background words were also very difficult to hear and process (as in Experiment 3 1SC/2SI); however, they were semantically processed as revealed by the semantic priming effect observed. As SC voices were more salient in Experiment 2 2SC/1SI than in Experiment 3 1SC/2SI, participants might have heard more related words and one could argue that semantic priming in our study in fact relied on the chance for participants to hear a SC word. Although this hypothesis is interesting, it does not seem sufficient to explain our results. Indeed, according to this hypothesis, a priming effect should have appeared also in Experiment 3 where participants, in line with previous results from the literature (Brungart, 2001; Hoen et al., 2007), recognized SC words presented in the post-test. Given the overall results of our experiments, we argue that if the ratio of SC/total voices was $<1/2$, participants heard some SC words, but, because of high cognitive load (i.e., intelligibility was low and they focused on the target item) these words were not processed sufficiently deeply to lead to semantic priming (a similar dissociation between lexical and semantic processing was also found by Kouider and Dupoux, 2005). However, if the ratio was $>1/2$, SC words were salient, and participants thus, processed them sufficiently to improve their performance. As this interpretation relies on the post-test effect which is quite small, although significant, more experiments should be performed. For example, using only SC voices and degrading intelligibility by adding noise or filtering the signal instead of adding SI voice would be a good way to test this hypothesis in future studies.

Our results suggest that the increased cognitive load necessary to reconstruct the degraded signal reduced available resources for higher-level processes. This claim is consistent with the Effortfulness Hypothesis (Rabbitt, 1968) that states degraded signals require allocating many cognitive resources to formal processes (i.e., orthographic or phonological), leaving less available cognitive resources to perform higher-level processes (e.g., lexical). It has been shown that hearing-impaired participants are less accurate at recalling previously heard final sentence words than their control peers (Pichora-Fuller et al., 1995). The underlying assumption is that for hearing-impaired participants, the auditory signal is highly degraded and therefore demands more cognitive resources to be formally (i.e., phonologically) processed. As studies usually use recall tasks of lists of unrelated words or digits (Surprenant, 1999; Murphy et al., 2000; Wingfield et al., 2005), verbal working memory only relies on the phonological loop (Baddeley and Hitch, 1974; Baddeley, 2000) and only involves formal processes. Our results showed that higher levels of processing such as semantic activation may be specifically impacted by signal degradation.

Recently, semantic and syntactic integration difficulties if a signal is degraded have been reported in the visual modality (Gao et al., 2011, 2012). Experiments conducted by Gao et al. (2011) using visual noise (i.e., pixel's brightness variation) showed that if participants allocate more resources to formal processes, semantic integration is affected. Indeed, after reading an entire text, participants were worse at recalling the main proposition in the noisy condition. Altogether, these findings suggest that the availability of cognitive resources is involved at various levels during language processing. Whereas previous studies have shown that noise impairs memory systems, our study provides evidence that semantic activation is linked to cognitive resources, independently of memory.

CONCLUSION

This study explored the semantic nature of informational masking in a cocktail party situation. The results of three behavioral studies reveal that the emergence of semantic priming effects relies on prime intelligibility and saliency. These findings question the assumption that signal degradation has no effect on speech processing if target signals can be recognized. The results reveal that high-level processes, such as semantic processing, might not be as automatic as previously thought but are subjected to the limits of cognitive resources. Our study also demonstrates how the cocktail party situation can be used to study the automaticity of linguistic processes.

ACKNOWLEDGMENTS

The first author is funded by a PhD grant from Rhône-Alpes region, France. This research was supported by a European Research Council grant to the SpiN project (no. 209234).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fnhum.2014.00878/abstract>

REFERENCES

- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2
- Baddeley, A., and Hitch, G. (1974). "Working Memory," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 8, ed G. H. Bower (New York, NY: Academic Press), 47–89.
- Boulenger, V., Hoen, M., Ferragne, E., Pellegrino, E., and Meunier, F. (2010). Real-time lexical competitions during speech-in-speech comprehension. *Speech Commun.* 52, 246–253. doi: 10.1016/j.specom.2009.11.002
- Bronkhorst, A. W. (2000). The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acustica* 86, 117–128.
- Brouwer, S., Van Engen, K., Calandruccio, L., and Bradlow, A. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *J. Acoust. Soc. Am.* 131, 1449–1463. doi: 10.1121/1.3675943
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696
- Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J. Acoust. Soc. Am.* 120, 4007–4018. doi: 10.1121/1.2363929
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple

- simultaneous talkers. *J. Acoust. Soc. Am.* 110, 2527–2538. doi: 10.1121/1.1408946
- Cherry, C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Collins, A., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407
- Collins, A., and Quillian, M. R. (1969). Retrieval time from semantic memory. *J. Verbal Learn. Verbal Behav.* 9, 240–247. doi: 10.1016/S0022-5371(69)80069-1
- Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *J. Acoust. Soc. Am.* 123, 414–427. doi: 10.1121/1.2804952
- Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature* 395, 597–600. doi: 10.1038/26967
- Donnenwerth-Nolan, S., Tanenhaus, M. K., and Seidenberg, M. S. (1981). Multiple code activation in word recognition: evidence from rhyme monitoring. *J. Exp. Psychol. Hum. Learn.* 7, 170–180. doi: 10.1037/0278-7393.7.3.170
- Drullman, R., and Bronkhorst, A. W. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *J. Acoust. Soc. Am.* 107, 2224–2235. doi: 10.1121/1.428503
- Dupoux, E., Kouider, S., and Mehler, J. (2003). Lexical access without attention? Explorations using dichotic priming. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 172–184. doi: 10.1037/0096-1523.29.1.172
- Durlach, N. (2006). Auditory masking: need for improved conceptual structure. *J. Acoust. Soc. Am.* 120, 1787–1790. doi: 10.1121/1.2335426
- Eich, E. (1984). Memory for unattended events: remembering with and without awareness. *Mem. Cognit.* 12, 105–111. doi: 10.3758/BF03198423
- Festen, J. M., and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.* 88, 1725–1736. doi: 10.1121/1.400247
- Forster, K. I., and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 680–698. doi: 10.1037/0278-7393.10.4.680
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J. Acoust. Soc. Am.* 115, 2246–2256. doi: 10.1121/1.1689343
- Gao, X., Levinthal, B. R., and Stine-Morrow, E. A. (2012). The effects of ageing and visual noise on conceptual integration during sentence reading. *Q. J. Exp. Psychol. (Hove)* 65, 1833–1847. doi: 10.1080/17470218.2012.674146
- Gao, X., Stine-Morrow, E. A., Noh, S. R., and Eskew, R. T. Jr. (2011). Visual noise disrupts conceptual integration in reading. *Psychon. Bull. Rev.* 18, 83–88. doi: 10.3758/s13423-010-0014-4
- Gautreau, A., Hoen, M., and Meunier, F. (2013). Let's all speak together! Exploring the masking effects of various languages on spoken word identification in multi-linguistic babble. *PLoS ONE* 8:e65668. doi: 10.1371/journal.pone.0065668
- Grainger, J., and Holcomb, P. J. (2009). Watching the word go by: on the time-course of component processes in visual word recognition. *Lang. Linguist. Compass* 3, 128–156. doi: 10.1111/j.1749-818X.2008.00121.x
- Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J. Acoust. Soc. Am.* 115, 833–843. doi: 10.1121/1.1639908
- Hoen, M., Meunier, F., Grataloup, C.-L., Pellegrino, F., Grimault, N., Perrin, F., et al. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Commun.* 49, 905–916. doi: 10.1016/j.specom.2007.05.008
- Hohlfeld, A., Sangals, J., and Sommer, W. (2004). Effects of additional tasks on language perception: an event-related brain potential investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1012–1025. doi: 10.1037/0278-7393.30.5.1012
- Hohlfeld, A., and Sommer, W. (2005). Semantic processing of unattended meaning is modulated by additional task load: evidence from electrophysiology. *Brain Res. Cogn. Brain Res.* 24, 500–512. doi: 10.1016/j.cogbrainres.2005.03.001
- Klauer, K. C., Eder, A. B., Greenwald, A. G., and Abrams, R. L. (2007). Priming of semantic classifications by novel subliminal prime words. *Conscious. Cogn.* 16, 63–83. doi: 10.1016/j.concog.2005.12.002
- Kouider, S., and Dehaene, S. (2009). Subliminal number priming within and across the visual and auditory modalities. *Exp. Psychol.* 56, 418–433. doi: 10.1027/1618-3169.56.6.418
- Kouider, S., and Dupoux, E. (2005). Subliminal speech priming. *Psychol. Sci.* 16, 617–625. doi: 10.1111/j.1467-9280.2005.01584.x
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 451–468. doi: 10.1037/0096-1523.21.3.451
- Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn. Sci.* 9, 75–82. doi: 10.1016/j.tics.2004.12.004
- Lewis, J. L. (1970). Semantic processing of unattended messages using dichotic listening. *J. Exp. Psychol.* 85, 225–228. doi: 10.1037/h0029518
- Marslen-Wilson, W., and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cogn. Psychol.* 10, 29–63. doi: 10.1016/0010-0285(78)90018-X
- Mattys, S. L., Brooks, J., and Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cogn. Psychol.* 59, 203–243. doi: 10.1016/j.cogpsych.2009.04.001
- Mattys, S. L., and Wiget, L. (2011). Effects of cognitive load on speech recognition. *J. Mem. Lang.* 65, 145–160. doi: 10.1016/j.jml.2011.04.004
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- McDermott, J. H. (2009). The cocktail party problem. *Curr. Biol.* 19, R1024–R1027. doi: 10.1016/j.cub.2009.09.005
- Meyer, D. E., and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *J. Exp. Psychol.* 90, 227–234. doi: 10.1037/h0031564
- Murphy, D. R., Craik, F. I., Li, K. Z., and Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychol. Aging* 15, 323–334. doi: 10.1037/0882-7974.15.2.323
- Naccache, L., and Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition* 80, 215–229. doi: 10.1016/S0010-0277(00)00139-6
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of Inhibitionless spreading activation limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–254. doi: 10.1037/0096-3445.106.3.226
- New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique* 101, 447–462. Available online at: <http://www.lexique.org>
- Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282
- Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Q. J. Exp. Psychol.* 20, 241–248. doi: 10.1080/14640746808400158
- Radeau, M. (1983). Semantic priming between spoken words in adults and children. *Can. J. Psychol.* 37, 547–556. doi: 10.1037/h0080756
- Radeau, M., Besson, M., Fonteneau, E., and Castro, S. L. (1998). Semantic, repetition and rime priming between spoken words: behavioral and electrophysiological evidence. *Biol. Psychol.* 48, 183–204. doi: 10.1016/S0301-0511(98)00012-X
- Schacter, D. L., and Church, B. A. (1992). Auditory priming: implicit and explicit memory for words and voices. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 915–930. doi: 10.1037/0278-7393.18.5.915
- Simpson, S. A., and Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *J. Acoust. Soc. Am.* 118, 2775–2778. doi: 10.1121/1.2062650
- Spruyt, A., De Houwer, J., Everaert, T., and Hermans, D. (2012). Unconscious semantic activation depends on feature-specific attention allocation. *Cognition* 122, 91–95. doi: 10.1016/j.cognition.2011.08.017
- Surprenant, A. (1999). The effect of noise on memory for spoken syllables. *Int. J. Psychol.* 34, 328–333. doi: 10.1080/002075999399648
- Van den Bussche, E., Van den Noortgate, W., and Reynvoet, B. (2009). Mechanisms of masked priming: a meta-analysis. *Psychol. Bull.* 135, 452–477. doi: 10.1037/a0015329
- Van Engen, K., and Bradlow, A. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *J. Acoust. Soc. Am.* 121, 519–526. doi: 10.1121/1.2400666
- Van Petten, C. (2014). “Selective attention, processing load, and semantics: insights from human electrophysiology,” in *Cognitive Electrophysiology of Attention*, ed G. R. Mangun (New York, NY: Harbound), 236–253.
- Wingfield, A., Tun, P. A., and McCoy, S. L. (2005). Hearing loss in Older Adulthood. What it is and how it interacts with cognitive performances. *Curr. Dir. Psychol. Sci.* 14, 144–148. doi: 10.1111/j.0963-7214.2005.00356.x

Wood, N. L., Stadler, M. A., and Cowan, N. (1997). Is there implicit memory without attention? A reexamination of task demands in Eich's (1984) procedure. *Mem. Cognit.* 25, 772–779. doi: 10.3758/BF03211320

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 March 2014; accepted: 12 October 2014; published online: 31 October 2014.

Citation: Dekerle M, Boulenger V, Hoen M and Meunier F (2014) Multi-talker background and semantic priming effect. *Front. Hum. Neurosci.* 8:878. doi: 10.3389/fnhum.2014.00878

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Dekerle, Boulenger, Hoen and Meunier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ADVANTAGES OF PUBLISHING IN FRONTIERS



FAST PUBLICATION

Average 90 days
from submission
to publication



COLLABORATIVE PEER-REVIEW

Designed to be rigorous –
yet also collaborative, fair and
constructive



RESEARCH NETWORK

Our network
increases readership
for your article



OPEN ACCESS

Articles are free to read,
for greatest visibility



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



GLOBAL SPREAD

Six million monthly
page views worldwide



COPYRIGHT TO AUTHORS

No limit to
article distribution
and re-use



IMPACT METRICS

Advanced metrics
track your
article's impact



SUPPORT

By our Swiss-based
editorial team